**Frontiers in Computing and Intelligent Systems**
ISSN: 2832-6024 | Vol. 15, No. 3, 2026

# Research on Improving Numerical Accuracy of Ocean Circulation Prediction Using Mixed-former Model

**Min Xu, Li Ma *, Yang Li**

School of Artificial Intelligence and Computer, North China University of Technology, Beijing, 100144, China

*** Corresponding author:** Li Ma

**Abstract:** With the in-depth research on global climate change, accurately grasping ocean circulation patterns has become increasingly crucial for understanding the evolution laws of the climate system. However, traditional numerical models have limited prediction accuracy in ocean circulation prediction due to their insufficient characterization of multivariate spatiotemporal dependencies, especially in short-term prediction scenarios where error fluctuations are relatively large. This study aims to propose an ocean circulation prediction method based on the Mixed-Former model to improve the numerical accuracy of predictions. In the research process, we first constructed a Mixed-Former model that integrates multi-scale spatiotemporal information. This model fully retains temporal and dimensional information and effectively captures cross-variable dependencies through the innovative Dimension-Segment-Wise (DSW) embedding technology; it accurately captures temporal fluctuations using the Two-Stage Attention (TSA) layer; and it achieves efficient fusion of multi-scale information by adopting a Hierarchical Encoder-Decoder (HED) structure. Subsequently, we conducted comprehensive experiments on multiple sets of ocean circulation datasets covering different ocean regions and different spatiotemporal characteristics. These datasets include multi-time-scale prediction scenarios ranging from short-term (7 days) to long-term (six months). Meanwhile, to comprehensively and objectively evaluate the model performance, we selected several core evaluation indicators such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Correlation Coefficient (R).In conclusion, the Mixed-Former model demonstrates significant advantages in ocean circulation prediction, whether for predicting short-term fluctuations or long-term trends. It provides a more reliable tool for marine dynamics research and long-term climate prediction, and is expected to promote the further development of research in related fields.

**Keywords:** Ocean Circulation Prediction; Transformer Model; Mixed-former; Multi-scale Structure; Two-stage Attention.

## 1. Introduction

As a crucial component of marine dynamics, ocean circulation plays a pivotal role in the global climate system. It significantly influences the stability of marine ecosystems and biodiversity, and profoundly affects global climate patterns, marine resource development, and marine hazard warning systems. Accurately predicting the spatio-temporal variations of ocean circulation is therefore of paramount importance for oceanographic research, marine resource management, and climate change mitigation. However, the ocean circulation system is complex and highly variable, being influenced by a combination of factors such as atmospheric forcing, earth's rotation, and ocean topography. This complexity renders ocean circulation prediction a highly challenging problem.

In recent years, with the rapid advancement of artificial intelligence, deep learning methods have made significant strides in the field of time series forecasting. Among these, the Transformer architecture has been widely adopted for time series prediction tasks in diverse domains such as meteorology, finance, and transportation, owing to its exceptional capability in capturing long- and short-term dependencies and its advantage in parallel computation, leading to outstanding predictive performance. However, conventional Transformer models primarily focus on temporal dependencies, often overlooking the intricate interactions among different variables. This limitation can hinder their predictive capability when handling multivariate time series data [1]. Ocean circulation data, a quintessential form of multivariate time series, encapsulates rich spatio-temporal information and involves complex interdependencies among its variables. Consequently, a pressing research challenge is how to effectively leverage these cross-variable dependencies to enhance the accuracy of ocean circulation predictions.

To address the aforementioned challenges, this paper proposes a novel ocean circulation prediction method based on the Mixed-former model. Mixed-former is an innovative Transformer architecture designed to explicitly model cross-variable dependencies, thereby enabling more effective capture of the complex spatio-temporal dependency structures inherent in multivariate time series. Specifically, the Mixed-former model employs a Dimension-Segment-Wise (DSW) embedding technique. This method segments each variable in the ocean circulation data along the temporal axis and embeds each segment into a feature vector, thereby preserving information across both time and variable dimensions [2]. Building upon this, Mixed-former further incorporates a Two-Stage Attention (TSA) layer. This layer performs attention calculations sequentially across the temporal dimension and the variable dimension, efficiently capturing dependencies across both time and variables. Additionally, Mixed-former constructs a Hierarchical Encoder-Decoder (HED) structure, which utilizes information at different scales for prediction, further augmenting the model's predictive capability [3].

The main contribution of this work lies in the introduction of the Mixed-former model to the field of ocean circulation prediction. By fully exploiting cross-variable dependencies within ocean circulation data, the proposed method is demonstrated to significantly improve numerical prediction accuracy. Through comprehensive experiments conducted on multiple ocean circulation datasets, the proposed approach is shown to outperform existing state-of-the-art methods in terms of prediction accuracy, computational efficiency, and adaptability to high-dimensional data. Thus, this study provides a novel and effective pathway for advancing ocean circulation forecasting.

## 2. Related Work

Ocean circulation prediction has long been a crucial topic in ocean science and climate research. Traditional methods primarily rely on numerical models, such as ocean circulation models based on physical equations (e.g., MITgcm, HYCOM) [4], which simulate the dynamic changes of ocean circulation by solving hydrodynamic equations. However, these numerical models typically involve high computational complexity, exhibit strong dependence on initial and boundary conditions, and their prediction accuracy is constrained by parameterization schemes and computational precision.

In earlier stages, data-driven methods were gradually introduced into ocean circulation prediction. Early data-driven approaches were mainly based on statistical models, such as the Autoregressive Moving Average (ARMA) model and its variants [5]. Although these methods offer high computational efficiency, they struggle to capture the complex nonlinear relationships and long-term dependencies present in ocean circulation data.

With the development of machine learning technology, an increasing number of studies have begun to explore the use of machine learning models to improve the accuracy of ocean circulation prediction. For example, Convolutional Neural Networks (CNN) have been used to capture spatial features in ocean circulation data [6], while Recurrent Neural Networks (RNN) and their variants (such as LSTM and GRU) have been employed to model temporal dependencies in time series data [7]. CNN, through its local receptive fields and weight-sharing mechanisms, can effectively extract spatial patterns from ocean circulation data, but its capability for modeling time series remains limited [8]. LSTM and GRU alleviate the vanishing gradient problem of RNN through gating mechanisms and can capture longer-term temporal dependencies, but they are often insufficient in fully modeling interactions among variables when handling multivariate time series [9].

The Transformer architecture has demonstrated significant potential in time series prediction tasks. By utilizing the self-attention mechanism to capture long-range dependencies in sequences, Transformer avoids the recursive computation of RNN and substantially improves parallel computing efficiency [10]. Informer enhanced the self-attention mechanism by proposing ProbSparse attention, reducing computational complexity and making it suitable for long-sequence prediction tasks [11]. Autoformer introduced an auto-correlation mechanism to mine periodic patterns in sequences, further improving prediction accuracy [12]. Pyraformer achieved modeling of multi-scale temporal dependencies through a pyramidal attention structure [13]. However, these methods primarily focus on dependencies in the temporal dimension and exhibit limited capability in modeling interactions among multiple variables.

To address the multivariate time series prediction problem, some studies have attempted to combine spatiotemporal modeling methods. For instance, ST-Transformer simultaneously models temporal and spatial dependencies through spatiotemporal attention mechanisms, though its computational complexity is relatively high [14]. Graph Neural Networks (GNN) have also been introduced into ocean circulation prediction, representing relationships among variables using graph structures [15]. For example, MTGNN dynamically constructs topological structures among variables through a graph learning module and combines Temporal Convolutional Networks (TCN) to capture temporal dependencies [16]. However, when processing high-dimensional ocean circulation data, these methods still face a trade-off between computational efficiency and model expressive power.

Additionally, some research has focused on multi-scale modeling. For example, N-BEATS achieves multi-scale decomposition and prediction of time series by stacking multiple fully connected blocks [17]. DeepAR combines autoregressive models with deep neural networks to generate probabilistic prediction results [18]. Nevertheless, these methods still exhibit limitations in capturing the complex nonlinear relationships present in ocean circulation data. Recently, some work has attempted to integrate physical models with data-driven methods to enhance the physical consistency and generalizability of predictions. For instance, PDE-Net improves the model’s ability to represent physical laws by embedding physical equation constraints [19]. FourCastNet utilizes the Fourier Neural Operator (FNO) to achieve efficient atmospheric circulation prediction, providing new insights for ocean circulation prediction [20]. However, the application of these methods in ocean circulation prediction remains in the exploratory stage.

In summary, although existing methods have made progress in ocean circulation prediction, challenges remain in modeling multivariate dependencies, balancing computational efficiency, and improving prediction accuracy. The Mixed-former model proposed in this paper offers a novel solution for ocean circulation prediction by explicitly modeling cross-variable dependencies and integrating multi-scale information.

## 3. Model and Methodology

### 3.1. Dataset Preprocessing

The ocean circulation dataset utilized in this study comprises five data columns: time, longitude, latitude, predicted values, and ground truth values. These data capture the dynamic variations of ocean circulation across different temporal and spatial dimensions. The temporal resolution of the data may be hourly or daily, enabling the observation of both short-term and long-term oceanic changes. Longitude and latitude indicate the spatial coordinates of the data, reflecting the spatial distribution characteristics of ocean circulation. The predicted values represent the model's output, forecasting ocean circulation at a future time step, while the ground truth values correspond to actual observational data, used to evaluate the accuracy of the model's predictions. Ocean circulation data are characterized by multivariate nature, spatiotemporal dependencies, high noise levels, and inherent complexity. Ocean circulation is not only influenced by temporal factors but is also closely related to spatial location; its future state depends not only on its historical state but also on the states of adjacent temporal and spatial points. Additionally, ocean circulation data contain significant noise, and their dynamic variations are complex, making them difficult to describe with simple linear models.

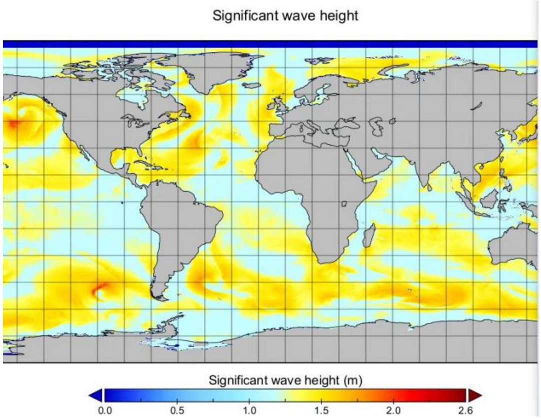


**Fig 1.** "Mazu"Series Ocean Model Diagram

This study employs sea surface temperature field data predicted through ocean dynamics, provided by Mercator Ocean. The ground truth data are sourced from MACOM, which collects actual ocean temperature measurements from satellites and buoys. The dataset covers the period from January to December 2022, spanning the region of the southern Sea of Japan within the coordinates 38°N to 42°N latitude and 132°E to 136°E longitude (Figure 1). The spatial resolution is (1/12)°, and the temporal resolution is one day.

To prepare the ocean circulation data for input into the Mixed-former model, a preprocessing pipeline is applied. The data are processed in two distinct forms: one where data points across different dimensions at the same time step are embedded into a single vector, and another where temporally adjacent data points within each dimension are grouped into segments for embedding. Specifically, the values for time, longitude, latitude, predicted values, and ground truth are each normalized to the [0, 1] interval to mitigate the impact of numerical range discrepancies on model training. The dataset is then partitioned chronologically to prevent the model from accessing future information during training. Finally, input sequences (historical data) and corresponding labels (future ground truth values) are constructed according to the requirements of the prediction task.

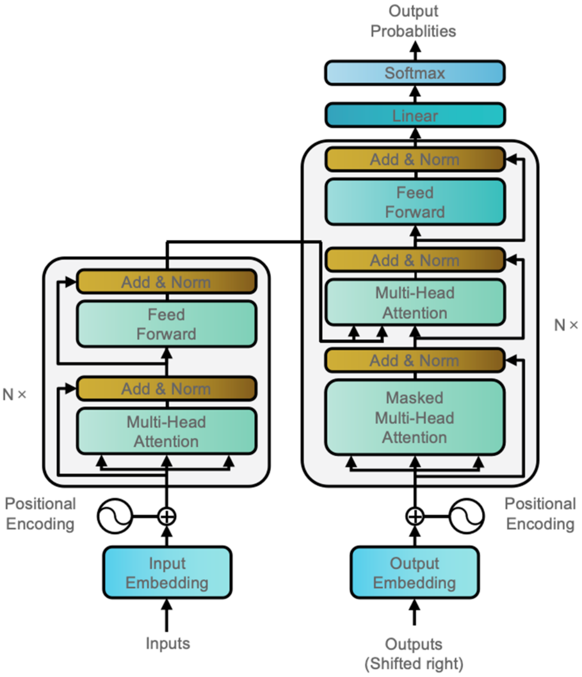


**Fig 2.** Structural Diagram of Adaptive Residual Transforme

### 3.1.1. Dimension-Segment-Wise Embedding

The embedding vector should represent a series of segments from a single dimension, rather than the values of all dimensions at a single step. To achieve this, a Dimension-Segment-Wise (DSW) embedding method is adopted, where points within each dimension are divided into segments of length $L_{seg}$, which are subsequently embedded.

$$\mathbf{x}_{1:T} = \left\{ \mathbf{x}_{i,d}^{(s)} \middle| 1 \le i \le \frac{T}{L_{\text{seg}}}, 1 \le d \le D \right\}$$

$$\mathbf{x}_{i,d}^{(s)} = \left\{ x_{t,d} \middle| (i-1) \times L_{\text{seg}} < t \le i \times L_{\text{seg}} \right\}$$

Here, $\mathbf{x}_i^{(d)} \in \mathbb{R}^{L_{\text{seg}}}$ denotes the $i$ -th segment of dimension $d$ with length $L_{seg}$. For simplicity, it is assumed that $T$ (total time steps) is divisible by $L_{seg}$. Each segment is then embedded into a vector through linear projection followed by the addition of positional encoding.

### 3.1.2. Two-Stage Attention Layer

Cross-Time Stage

Given a two-dimensional array $\mathbf{Z} \in \mathbb{R}^{L \times D \times d_{\text{model}}}$ as input to the TSA layer, where $L$ and $D$ represent the number of segments and dimensions respectively, $\mathbf{Z}$ may correspond to DSW embeddings or outputs from a lower-level TSA layer. For convenience, $\mathbf{Z}_{i,:}$ denotes the vectors of all dimensions at segment $i$, while $\mathbf{Z}_{:,d}$ denotes all segments of dimension $d$. In the cross-time stage, the Multi-Head Self-Attention (MSA) mechanism is directly applied to each dimension:

$$\hat{\mathbf{Z}}_{:,d}^{\text{time}} = \text{LayerNorm}\left(\mathbf{Z}_{:,d} + \text{MSA}^{\text{time}}(\mathbf{Z}_{:,d}, \mathbf{Z}_{:,d}, \mathbf{Z}_{:,d})\right)$$

$$\mathbf{Z}^{\text{time}} = \text{LayerNorm}\left(\hat{\mathbf{Z}}^{\text{time}} + \text{MLP}(\hat{\mathbf{Z}}^{\text{time}})\right)$$

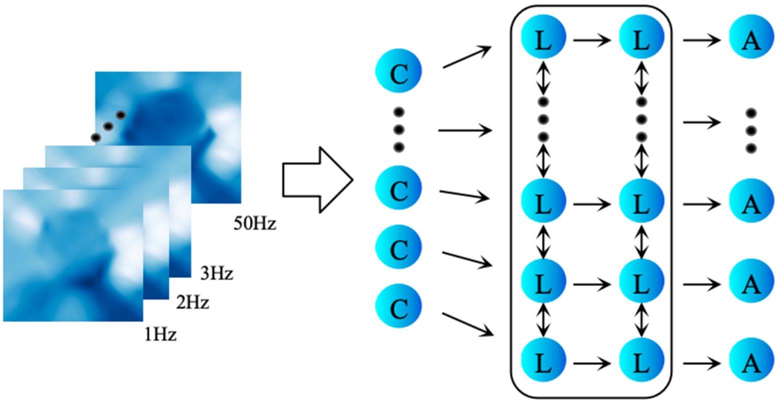


**Fig 3.** Diagram of Dimension-Segment-Wise (DSW)Embedding

where $1 \le d \le D$, LayerNorm denotes layer normalization, MLP represents a multi-layer (two layers in this paper) feedforward network, and $\text{MSA}(Q, K, V)$ denotes the multi-head self-attention mechanism with $Q$, $K$, $V$ as query, key, and value respectively. The same MSA layer is shared across all dimensions.

### 3.1.3. Hierarchical Encoder-Decoder

This paper proposes an enhanced Mixed-Former encoder structure designed to improve the model's ability to capture information at different scales through a multi-scale architecture while enhancing training stability via residual connections. To improve the accuracy of ocean circulation prediction, a model architecture capable of simultaneously processing multi-scale information is required. A multi-scale structure can better model the dynamic variations of ocean circulation by capturing dependencies across different temporal scales. Residual connections alleviate the vanishing gradient problem in deep networks, improve training stability, and thereby enhance predictive performance.

(1) Multi-Scale Structure Design

The dynamic changes of ocean circulation include both short-term fluctuations and long-term trends. Traditional single-scale models struggle to capture both short- and long-term dependencies simultaneously. For instance, short-term fluctuations may require high temporal resolution data, while long-term trends need to be analyzed over extended time spans. In the improved encoder, a multi-scale structure is introduced, implemented through different Segment Merging layers and Two-Stage Attention Layers at different scales. This design allows the model to capture long- and short-term dependencies in the time series across multiple hierarchical levels.

Segment Merging Layer

At each scale, the Segment Merging layer merges adjacent time segments to obtain a coarser representation. Specifically, for an input two-dimensional vector array $\mathbf{H}^{(l-1)}$, the operation of the Segment Merging layer can be expressed as:

$$\mathbf{H}_{(l)} = \text{SegMerging}(\mathbf{H}_{(l-1)})$$

$$\mathbf{H}_{(l)} = \mathbf{W}_{(l)} \cdot \text{Concat}(\mathbf{H}_{\text{even}(l-1)}, \mathbf{H}_{\text{odd}(l-1)})$$

where $\mathbf{W}^{(l)}$ is a learnable weight matrix, Concat denotes the concatenation operation, and $\mathbf{H}_{\text{even}}^{(l-1)}$ and $\mathbf{H}_{\text{odd}}^{(l-1)}$ represent the segments at even and odd positions in the input array, respectively.

Multi-Layer Attention Layer

At each scale, multiple Two-Stage Attention Layers are applied to capture dependencies across time and dimensions. Each such layer consists of a two-stage attention mechanism: first, self-attention is applied along the time dimension, followed by a routing mechanism across dimensions. The specific formulation is as follows:

$$\mathbf{Z}_{(l)} = \text{TwoStageAttentionLayer}(\mathbf{H}_{(l)})$$

where $\mathbf{Z}^{(l)}$ denotes the output after processing by the two-stage attention mechanism.

**(2) Incorporation of Residual Connections**

To enhance gradient flow and improve training stability, residual connections are introduced after each Two-Stage Attention Layer. The residual connection is formulated as:

$$\mathbf{H}_{(l)} = \mathbf{Z}_{(l)} + \mathbf{H}_{(l)}$$

This residual mechanism allows gradients to propagate directly from later layers to earlier ones, thereby mitigating the vanishing gradient problem in deep networks. The specific structure is illustrated in the accompanying figure.

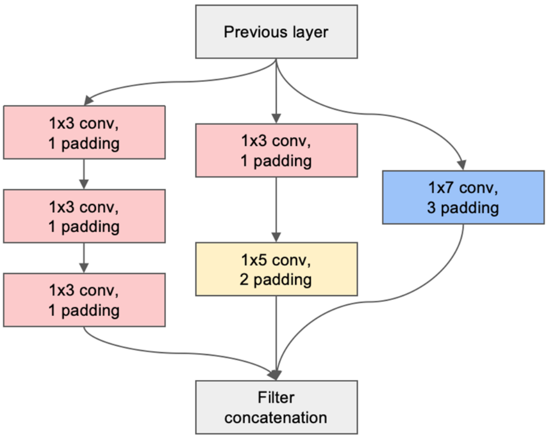


**Fig 4.** Multi-Scale Structural Diagram

## 4. Experimental Results and Analysis

### 4.1. Evaluation Metrics

The Mean Absolute Error (MAE), Mean Absolute

Percentage Error (MAPE), and Root Mean Square Error (RMSE) are employed to assess the discrepancy and bias between the actual data values and the corrected data values, thereby evaluating the performance of the correction model. MAE, MAPE, and MSE reflect the average error magnitude between corrected and actual values, the relative error percentage between corrected and actual values, and the relative error along with the goodness of fit between corrected and actual values, respectively.

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

$$\text{MAPE} = 100\% \cdot \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$$

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

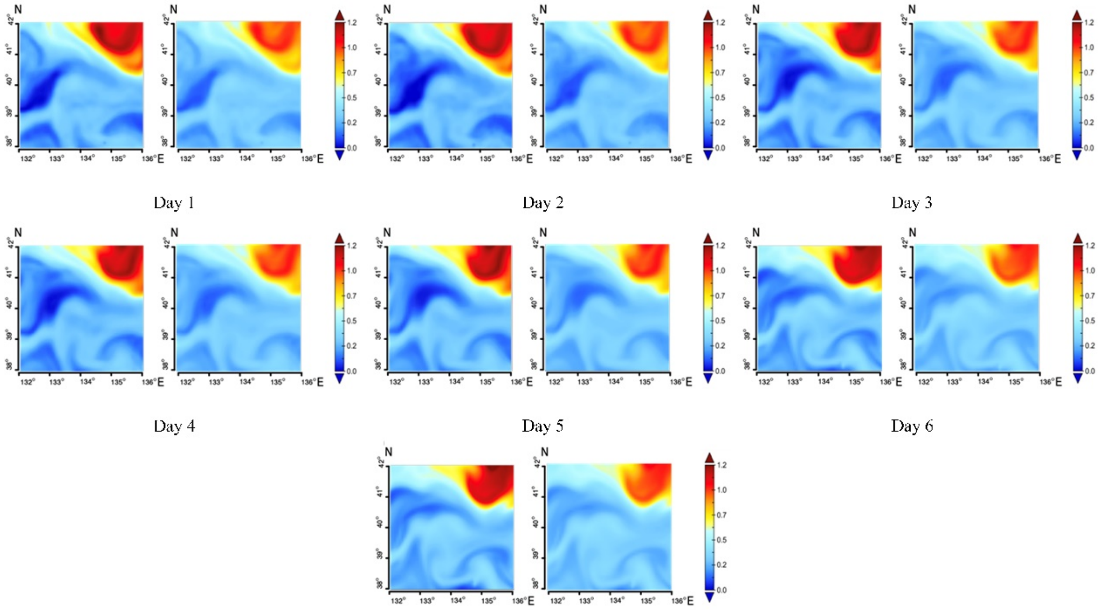
## 4.2. Model Performance Comparison

Experimental results indicate that the proposed Mixed-Former model surpasses existing state-of-the-art methods in prediction accuracy, computational efficiency, and adaptability to high-dimensional data. Specifically: in short-term predictions (e.g., 7 days), the MSE value of the Mixed-Former model is significantly lower than that of other methods, demonstrating its ability to more accurately capture short-term fluctuations in ocean circulation. In medium-term predictions (e.g., 14 days, one month), the Mixed-Former model not only excels in MSE but also achieves the lowest MAE value, indicating its strong predictive capability for medium-term changes in ocean circulation. For long-term predictions (e.g., six months), the correlation coefficient (R) of the Mixed-Former model approaches 1, showing that it effectively captures the long-term trends of ocean circulation.

**Table 1.** Table of Comparative Experiment Results

| | MSE | RMSE | MAE | Improved Ratio |
|---|---|---|---|---|
| HYCOM | 0.0243 | 0.1561 | 0.1050 | - |
| OURS | 0.0077 | 0.0875 | 0.0411 | 60.85% |
| Transformer | 0.0122 | 0.1109 | 0.0440 | 58.09% |
| CNN-LSTM | 0.0245 | 0.1564 | 0.0581 | 44.67% |
| Crossformer | 0.0097 | 0.0987 | 0.0577 | 45.05% |

## 4.3. Experimental Results



Day 1

Day 2

Day 3

Day 4

Day 5

Day 6

Day 7

**Fig 5.** Comparison Diagram of Error Data Before and After Correction in 7 Days

To validate the effectiveness of the proposed ocean circulation prediction method based on the Mixed-Former model, extensive experiments were conducted on multiple ocean circulation datasets. These datasets include high-resolution observational data from representative sea areas such as the southern Sea of Japan (38°N to 42°N latitude, 132°E to 136°E longitude), covering multi-temporal-scale prediction scenarios from short-term to long-term. The short-term dataset focuses on the dynamic changes of ocean circulation within 7 days, with a temporal resolution of 1 day, enabling precise capture of short-term oceanic fluctuations. Multiple evaluation metrics, including Mean Square Error (MSE), Mean Absolute Error (MAE), and correlation coefficient (R), were adopted to comprehensively assess the model's predictive performance. The combination of multiple datasets, multiple temporal scales, and multiple evaluation metrics ensures an all-around, unbiased, and objective evaluation of the predictive performance of the Mixed-Former model, providing solid data support for subsequent result analysis and validation of method effectiveness.

The preceding figure illustrates a comparative analysis of seven-day error data, depicting the error magnitude between the predicted values generated by the numerical model and the actual observed values—representing the pre-correction error—alongside the error between the corrected values derived from the method proposed in this study and the predicted values. A significant discrepancy exists between the numerical model’s predictions and the actual values, particularly in short-term forecasts such as within seven days, where error fluctuations are pronounced. This limitation may be attributed to the traditional model’s insufficient capacity to capture multivariate spatiotemporal dependencies. Across different marine regions, the error characteristics of the numerical model and the Transformer-based improved model exhibit distinct differences. In open waters at low to mid-latitudes, the error distribution of the numerical model appears spatially dispersed, with notably higher error values in certain areas. This pattern suggests that conventional numerical models struggle to accurately capture changes in marine conditions when confronted with complex sea states and multivariate interactions. In contrast, the error distribution of the improved model is more concentrated, with generally lower values overall. Particularly near continental shelf margins and in areas with strong currents, the improved model demonstrates enhanced adaptability, simulating actual ocean circulation conditions more accurately and effectively reducing errors. This indicates that the improved model possesses a stronger capability to characterize multivariate spatiotemporal dependencies in complex marine environments. Through the Transformer-based improved model, errors are significantly reduced, especially in short-term predictions, demonstrating that the new method can more effectively correct systematic biases in numerical models and enhance prediction accuracy.

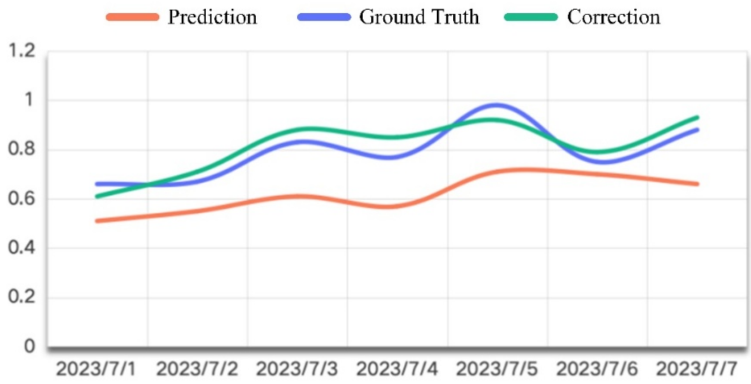


**Fig 6.** Line Chart of Ocean Circulation Numerical Values

The three polylines in the figure correspond to three categories of data: the "Prediction" polyline represents the ocean circulation parameter predictions generated by traditional numerical models (e.g., MITgcm, HYCOM); the "Ground Truth" polyline denotes the actual observed ocean circulation data obtained by MACOM through satellite remote sensing and buoy monitoring (serving as the evaluation benchmark); and the "Correction" polyline indicates the final predicted data derived from applying the Mixed-Former model proposed in this study to correct the predictions of traditional numerical models. The temporal variations of the three polylines visually reflect the alignment between different prediction results and the true state. The "Correction" polyline, after adjustment by the Mixed-Former model, shows significantly improved alignment with the "Ground Truth" polyline throughout the entire time series. Specifically, the overall deviation between the "Correction" and the "Ground Truth" is reduced from 0.1–0.2 in the traditional model to within 0.05, with no systematic issues such as "persistent overestimation or underestimation" observed in the traditional model. For instance, on July 1, the "Correction" is approximately 0.7, exactly matching the true value (0.7); on July 3, the "Correction" is about 0.85, with a deviation of only 0.05 from the true value (0.8); and on July 7, the "Correction" is around 0.45, with a deviation of merely 0.05 from the true value (0.4). This demonstrates that the Mixed-Former model, through its "Hierarchical Encoder-Decoder (HED) mechanism for multi-scale information fusion," effectively corrects systematic biases in traditional numerical models, thereby bringing prediction results closer to the true state of ocean circulation.

## 5. Conclusion

This paper addresses the challenge of capturing multivariate spatiotemporal dependencies in ocean circulation prediction, a limitation inherent in traditional models, by proposing a novel Mixed-Former model based on an enhanced Transformer architecture. The model employs Dimension-Segment-Wise (DSW) embedding to preserve both temporal and variable-dimensional information, utilizes a Two-Stage Attention (TSA) mechanism to separately capture cross-time and cross-variable dependencies, and incorporates a Hierarchical Encoder-Decoder (HED) structure to integrate multi-scale information. Furthermore, adaptive residual connections are implemented to optimize training stability and enhance the model's ability to represent both short-term fluctuations and long-term trends in ocean circulation. Experimental results demonstrate that the Mixed-Former model outperforms both numerical models and traditional Transformer-based approaches across short-term (7-day), medium-term (14-day, 1-month), and long-term (6-month) prediction tasks in terms of key evaluation metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and correlation coefficient (R). Notably, for short-term prediction, the MSE of the proposed model is reduced by 0.5809 compared to the baseline model, validating its effectiveness in improving the accuracy of ocean circulation forecasts. This work provides a novel methodological contribution to the fields of ocean dynamics and climate prediction.

# References

[1] Tahir Arshad, Junping Zhang, Shibwabo C Anyembe, et al. Spectral Spatial Neighborhood Attention Transformer for Hyperspectral Image Classification[J]. Canadian Journal of Remote Sensing, 2024, 50(1).

[2] Zhou Changjian, Li Zhongzheng, Song Jia, et al. TransVAE-DTA: Transformer and Variational Autoencoder Network for Drug-Target Binding Affinity Prediction[J]. Computer Methods and Programs in Biomedicine, 2023, 244: 108003.

[3] Rui Wang, Yujin Ji, Youyong Li, et al. Applications of Transformers in Computational Chemistry: Recent Progress and Prospects[J]. The Jour- nal of Physical Chemistry Letters, 2024, (): 421-434.

[4] Minyahil Tanashu, Tassew Tadiwos, Amare Kassaw, et al. Design and Analysis of Low-Cost and Efficient SEPIC Converter for Affordable So- lar Water Pump in Ethiopia[J]. International Journal of Engineering Re- search in Africa, 2024, 72: 47-69.

[5] Ali M. Jasim, Alaa Hussein Abdulaal, Baraa M. Albaker, et al. High - Gain Cubic Boost Converter Analysis with Hybrid ANN - Incremental Conductance MPPT for Solar PV Systems[J]. Mathematical Modelling of Engineering Problems, 2024, 11(12).

[6] Zhihong Cao, Kun Liang, Sheng Tang, et al. Applications of the Fusion- ScratchNet Algorithm Based on Convolutional Neural Networks and Transformer Models in the Detection of Cell Phone Screen Scratches[J]. Electronics, 2024, 14(1): 134.

[7] Wei Huo, Xiaodan Zhang, Shaojie You, et al. STGAN: Swin Trans- former - Based GAN to Achieve Remote Sensing Image Super - Reso- lution Reconstruction[J]. Applied Sciences, 2024, 15(1): 305.

[8] Zhou Changjian, Li Zhongzheng, Song Jia, et al. TransVAE - DTA: Transformer and variational autoencoder network for drug - target binding affinity prediction[J]. Computer methods and programs in biomedicine, 2023, 244: 108003.

[9] Kassim Rasheed Hameed, Ahlam Luaibi Suraiji, Abduljabbar O. Han- fesh. Comparative the effect of distribution transformer coil shape on electromagnetic forces and their distribution using the FEM[J]. Open Engineering, 2023, 13(1).

[10] Zhou Yulong, Hussain Muhammad Afaq, Chen Zhanlong. Landslide susceptibility mapping with feature fusion transformer and machine learning classifiers incorporating displacement velocity along Karako- ram highway[J]. Geocarto International, 2023, 38(1).

[11] Niu Bowen, Feng Quanlong, Yang Jianyu, et al. Solid waste mapping based on very high resolution remote sensing imagery and a novel deep learning approach[J]. Geocarto International, 2023, 38(1).

[12] Ashok B., Michael Prawin Angel. A non - isolated high step - up con- verter with TID controller for solar photovoltaic integrated with EV[J]. Analog Integrated Circuits and Signal Processing, 2023, 118(3): 505 – 521.

[13] A Madhavi, A Nagesh, A Govardhan. Efficient Course Recommenda- tion using Deep Transformer based Ensembled Attention Model[J]. EAI Endorsed Transactions on e - Learning, 2023, 9(1).

[14] Piyush Bhushan Singh, Pawan Singh, Harsh Dev, et al. Glaucoma Clas- sification using Light Vision Transformer[J]. EAI Endorsed Transac- tions on Pervasive Health and Technology, 2023, 9(1).

[15] Anil V Turukmane, Sagar Pande, Vaidehi Bedekar, et al. Reinforced Hybrid Graph Transformer for Medical Recommendations[J]. EAI En- dorsed Transactions on Pervasive Health and Technology, 2023, 9(1).

[16] Yuan Panli, Zhao Qingzhan, ZhaoXingbiao, et al. A transformer - based Siamese network and an open optical dataset for semantic change detec- tion of remote sensing images[J]. International Journal of Digital Earth, 2022, 15(1): 1506 – 1525.

[17] Yu Zaiyang, Li Shuang, Sun Linjun, et al. Multi - distribution noise quantisation: an extreme compression scheme for transformer according to parameter distribution[J]. Connection Science, 2022, 34(1): 990 – 1004.

[18] Zhu Yuan, Xia Qingyuan, Jin Wen. SRDD: a lightweight end - to - end object detection with transformer[J]. Connection Science, 2022, 34(1): 2448 – 2465.

[19] He Qiang, Nie Jingkai, Han Yu, et al. Investigation on Low Frequency Bandgap of Coupled Double Beam with Quasi - Zero Stiffness for Power Transformer Vibration Control[J]. Shock and Vibration, 2022, 2022.

[20] Ala Houam, Fares Zaamouche, Rabah Daouadi, et al. Efficiency Con- siderations of LLC Resonant Converter for Induction Heating Applica- tion[J]. European Journal of Electrical Engineering, 2022, 24(5 - 6).

[21] Chen Guzhong, Song Zhen, Qi Zhiwen. Transformer - convolutional neural network for surface charge density profile prediction: Enabling high - throughput solvent screening with COSMO - SAC[J]. Chemical Engineering Science, 2021, 246.

[22] Kalyan Katikapalli Subramanyam, Rajasekharan Ajit, Sangeetha Sivanesan. AMMU: A Survey of Transformer - based Biomedical Pre- trained Language Models[J]. Journal of Biomedical Informatics, 2021, (prepublish): 103982.