NMEFC, MNR, CHINA | liying.wan@nmefc.cn

High spatiotemporal resolution ocean temperature reconstruction and forecasting based on artificial intelligence

Liying Wan and Jiawei Jiang

April 2026

CONTENTS

Solving partial differential equations in numerical simulation is difficult and the calculation is complex.

On-site observation and sampling are costly, difficult to obtain, and have a sparse spatio-temporal distribution.

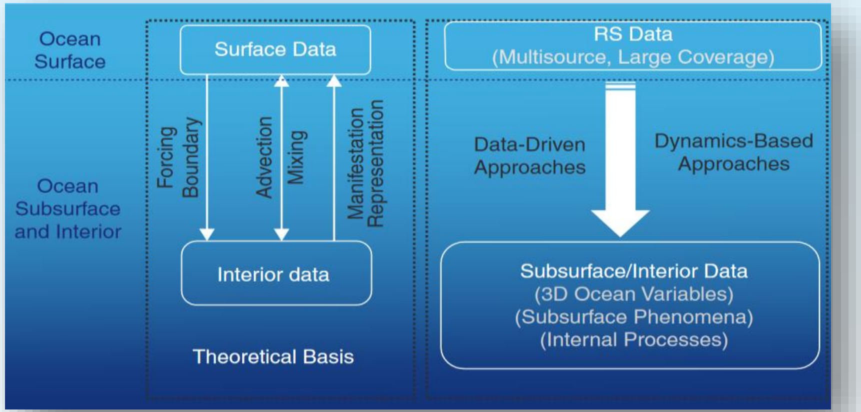
(Fox-Kemper et al., 2019; Bi et al., 2023)

AI

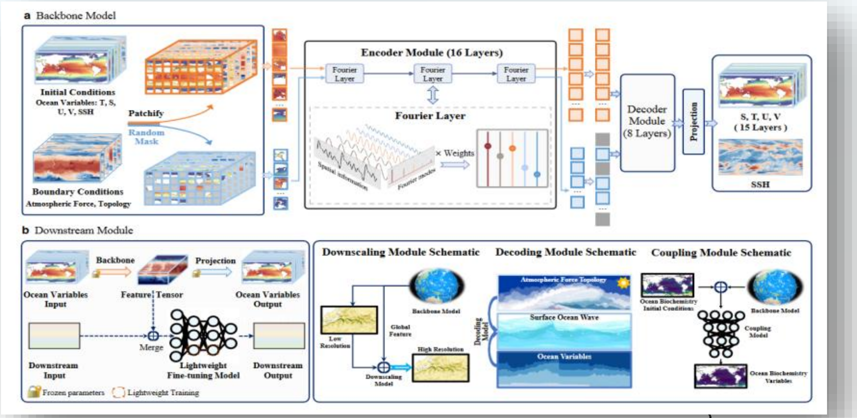
Satellite remote sensing inversion of the 3D ocean structure.

Large model for 3D structure prediction of the ocean

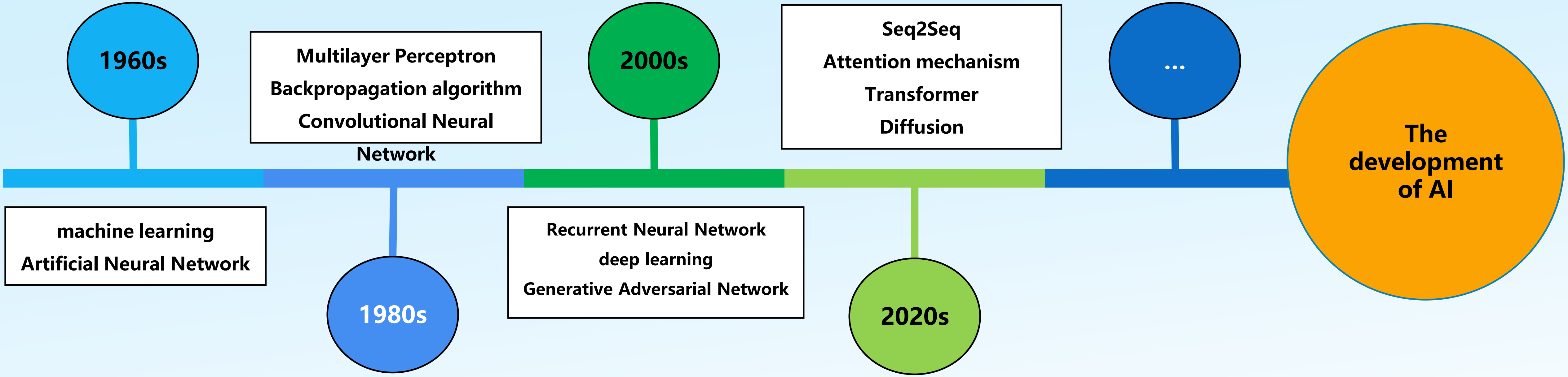
Real photos of the 2501 voyage of the National Natural Science Foundation of China



(Meng and Yan, 2022)



(Xiong et al.,



Reconstruction

1、**Single-point inversion（machine learning，ANN）** [B, C] (Ali et al., 2004; Smith et al., 2016; Lu et al., 2019)

2、**Regional inversion (CNN and its variants)** [B, C, H, W] （Meng et al., 2022; Xie et al., 2022）

3、**Time series (RNNS and their variants)** [B, C, T] （Su et al., 2021a, 2021b）

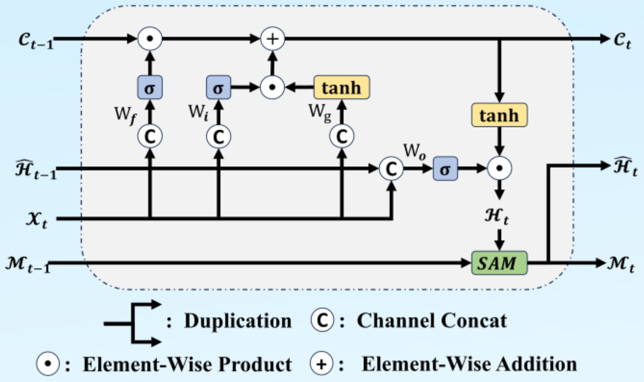
B represents the batch processing quantity. C is the data channel. H and W represent the length and width of the image, and T is the time step

Prediction

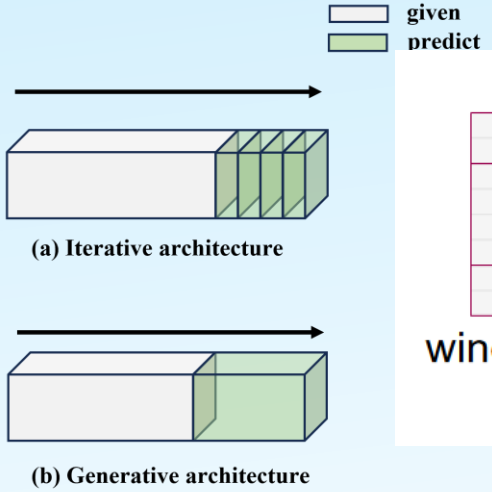
1、**Operational forecasting model（e.g. FOAM, PSY4）** （Blockley et al., 2014; Lellouche et al., 2018)

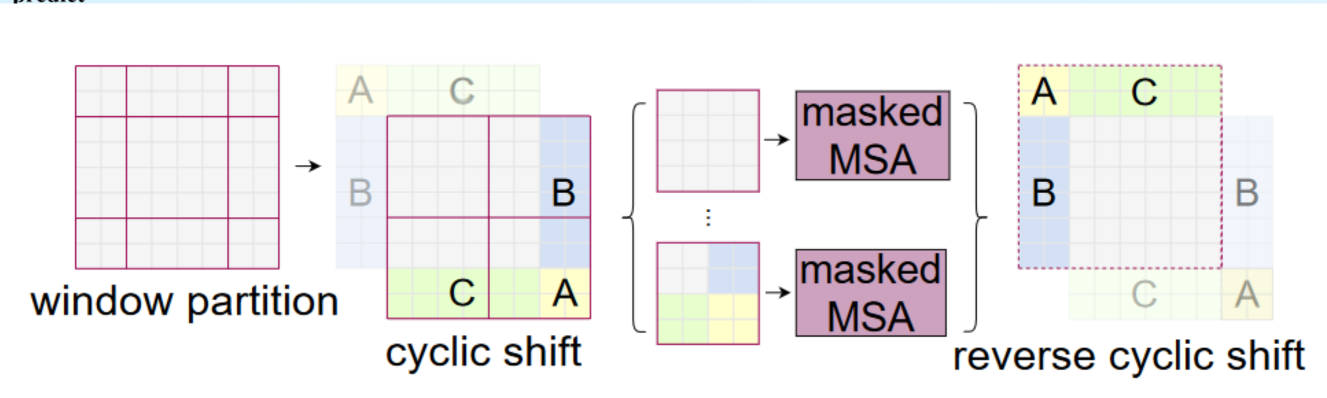
2、**Large Artificial Intelligence Model for Ocean Forecasting（e.g. AI-GOMS, XiHe）** （Xiong et al., 2023; Wang et al., 2024）

Research on the Earth system is related to time and space, so the spatiotemporal sequence [B, T, C, H, W] should be considered.



SA-ConvLSTM





Shifted Window

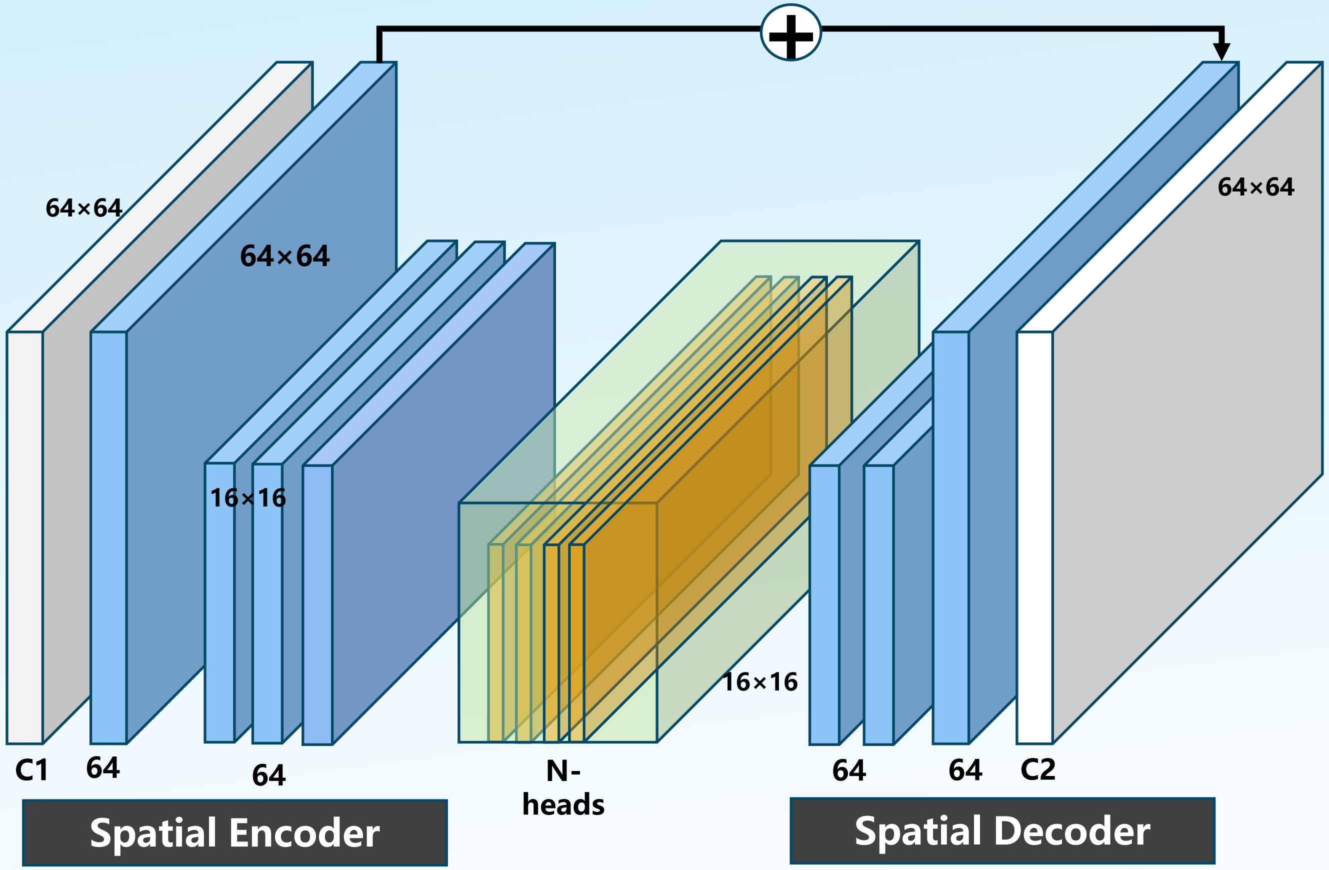
The RNN framework cannot capture long-term spatial memory very well (Lin et al. 2020)The results have the problem of error accumulation (Y. Wang et al. 2022; Tan et al. 2023)

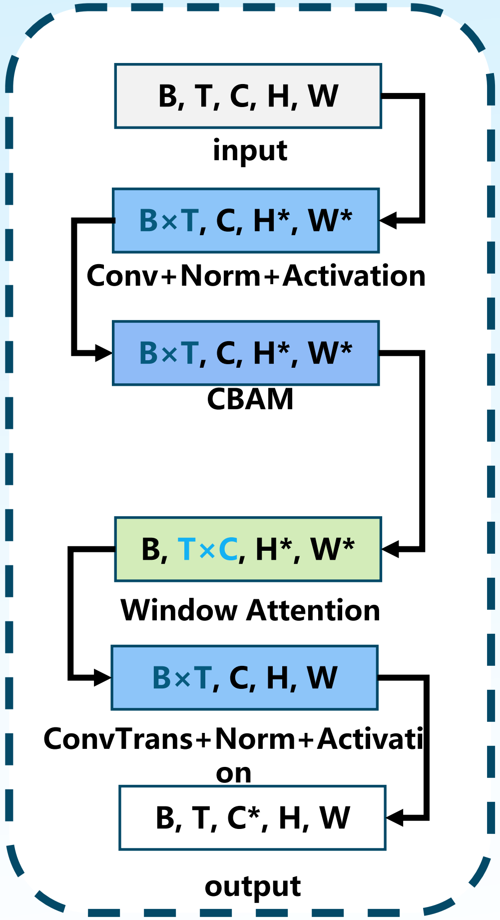
Attention（Transformer） + Generative forecasting（Informer） (H. Zhou et al. 2021; Lim et al. 2021)

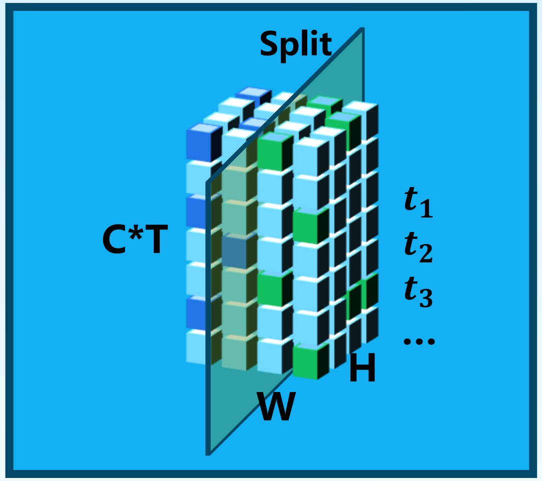
The global attention mechanism has a square-level complexity. As the image resolution increases and the time dimension is introduced, it becomes even more difficult to train

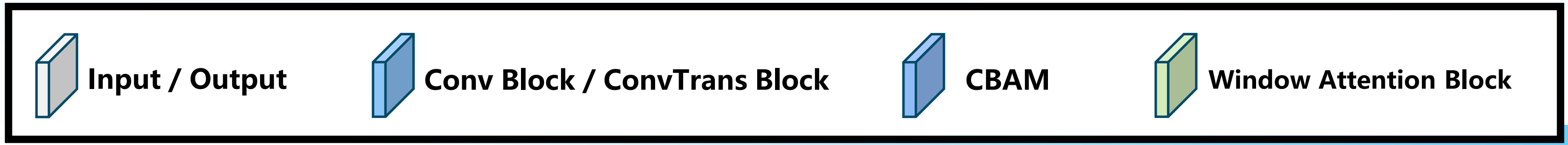
Shifted window mechanism（Swin Trans） + Spatio-temporal aggregation method（SimVP） (Liu et al., 2021; Tan et al., 2022)

Spatiotemporal Window Ocean Model (SWO)

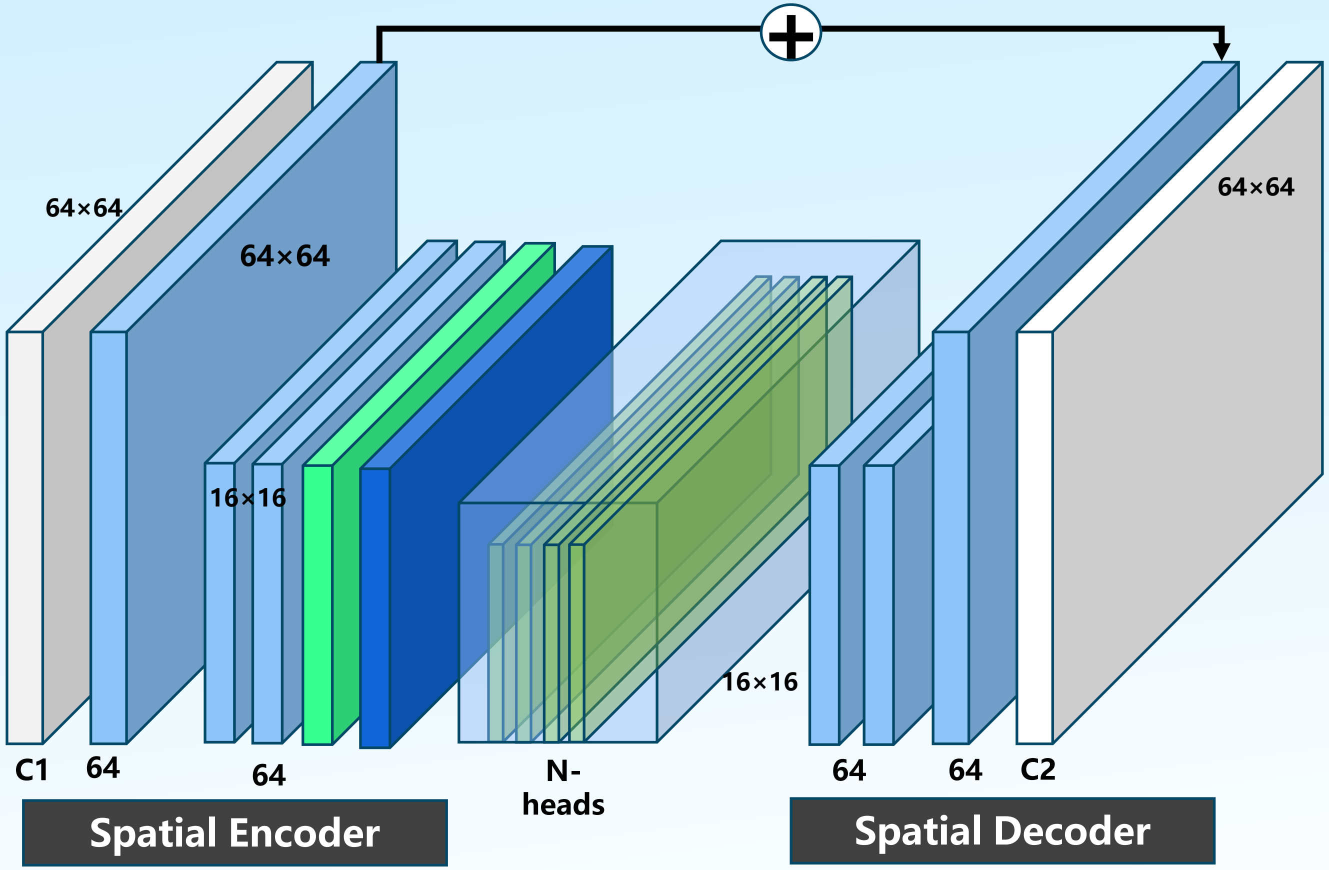


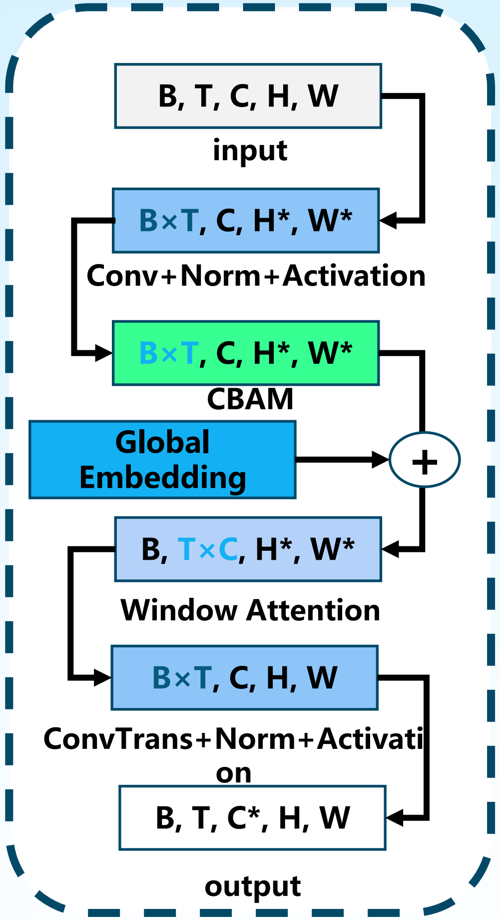


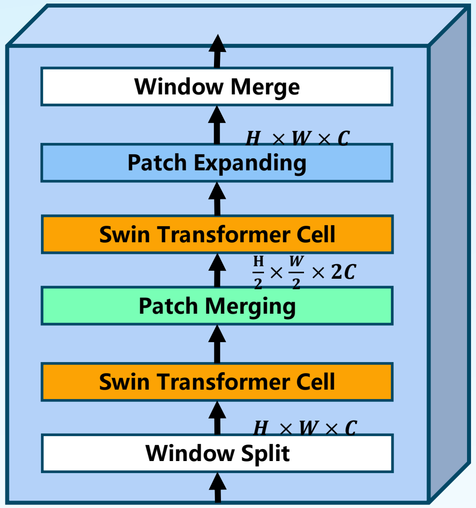


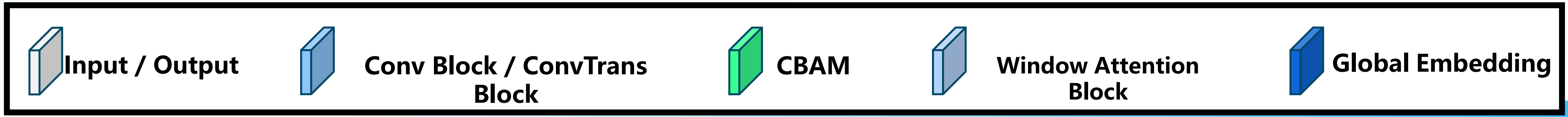


Multi-scale Residual Spatiotemporal Window Ocean Model (MSWO)

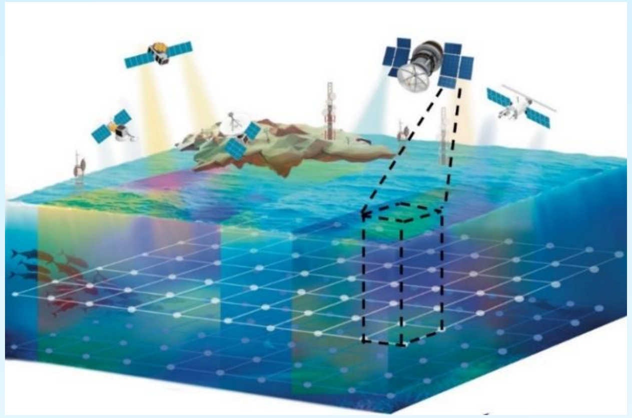




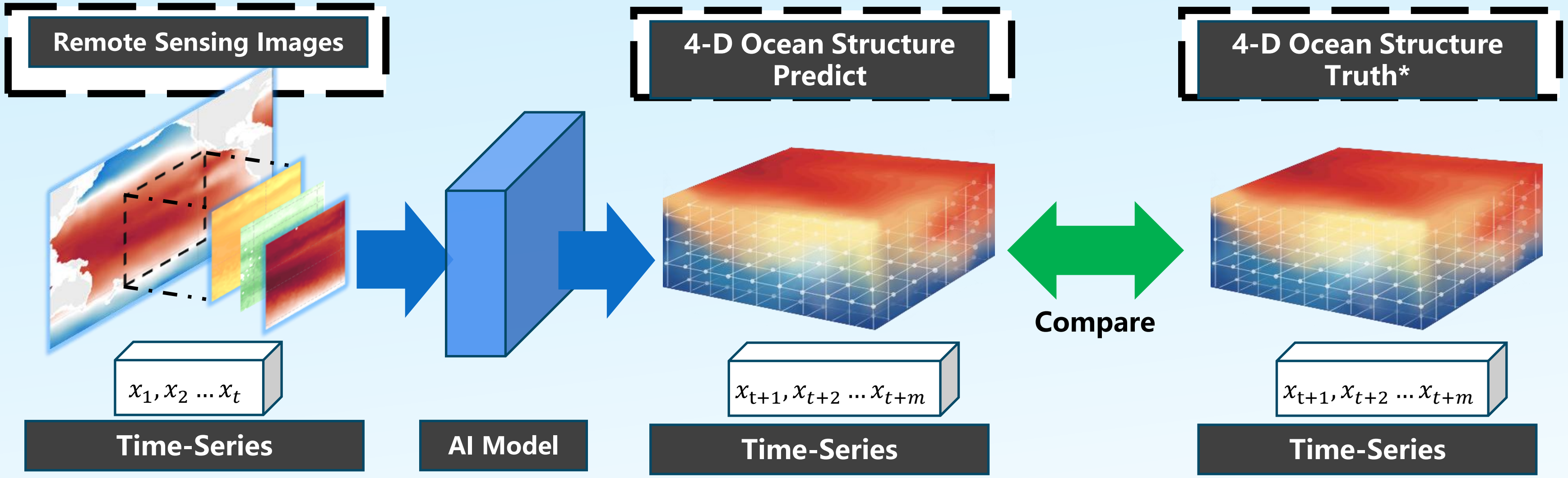


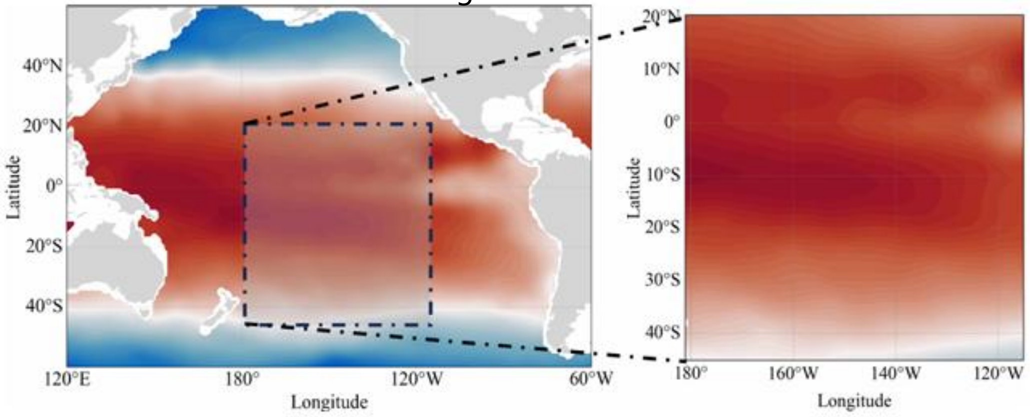


Construct a 3D temperature structure inversion model for the Pacific region based on spatio-temporal AI models



Ocean structure reconstructed by satellite remote sensing





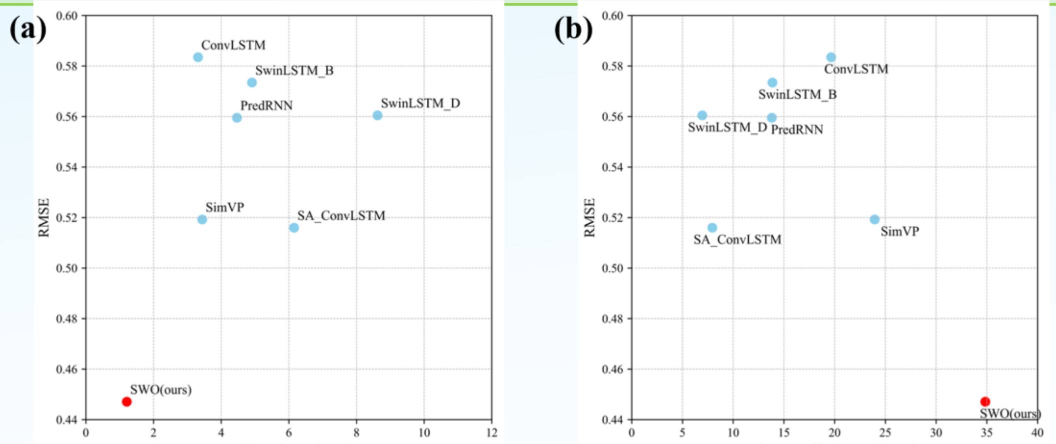
Study area (44.5°S~18.5°N, 116.5°W~179.5°W)

Parameter name	value
Time resolution	Monthly
Spatial resolution	1°
Dataset interval	2011-2019
Source	Argo
The proportion of training, validation, and test datasets	7: 1: 1
Framework	Pytorch
Loss	MSE

Parameter name	value
Optimizer	Adam
Batch Size	3
Epoch	1000
Early Stop	50
Learning Rate	0.001
GPU	Nvidia Tesla V100

Ablation experiment

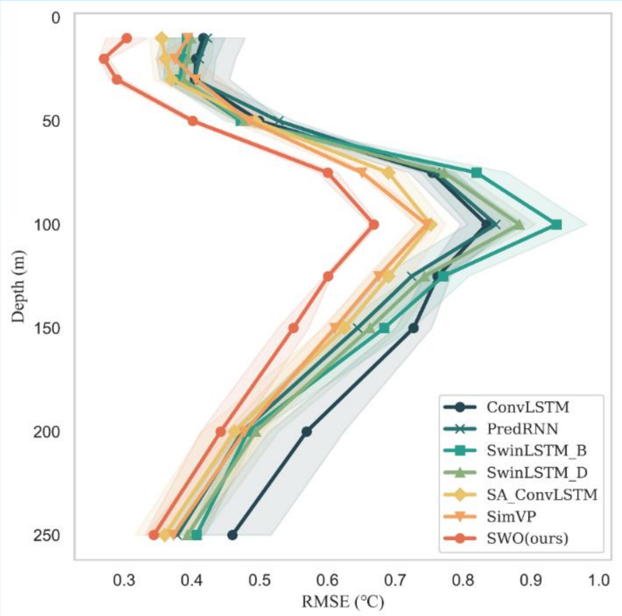
No.	CBAM	Window size	P.E.	Shifted	RMSE (°C) ↓	R^2 ↑
1	×	2	Rel	×	0.4985 ± 0.0164	0.9883 ± 0.0007
2	×	4	Rel	×	0.4563 ± 0.0224	0.9901 ± 0.0009
3	×	8	Rel	×	0.4897 ± 0.0181	0.9886 ± 0.0007
4	√	4	No	×	0.5007 ± 0.0200	0.9882 ± 0.0009
5	√	4	Abs	×	0.4904 ± 0.0274	0.9887 ± 0.0013
6	√	4	Rel	√	0.4719 ± 0.0129	0.9891 ± 0.0004
7	**√**	**4**	**Rel**	×	**0.4470 ± 0.0146**	**0.9905 ± 0.0006**



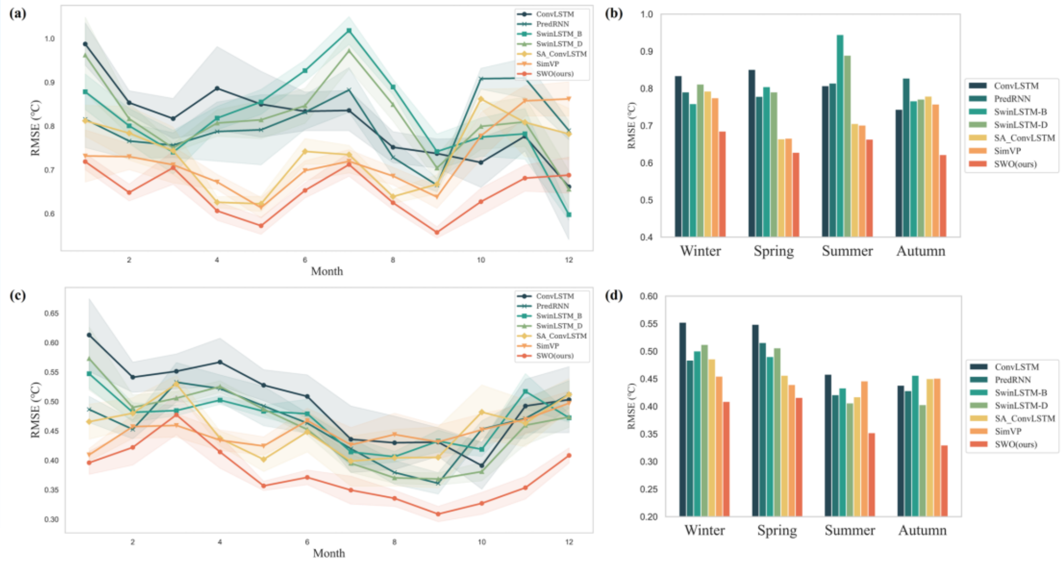
(a) The relationship between model training time (s/epoch) and RMSE.
(b) The relationship between model inference efficiency (frames/s) and RMSE.

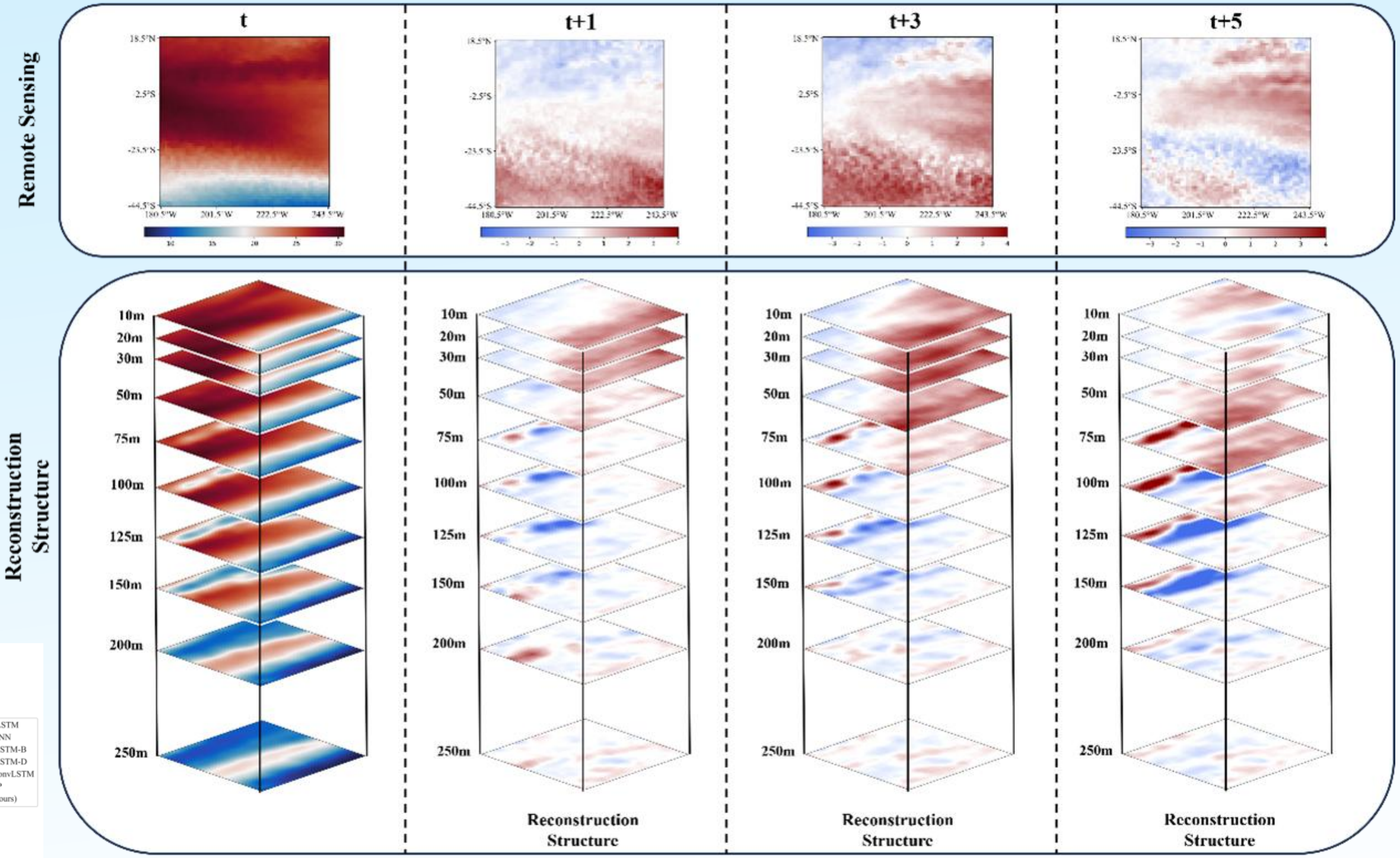
Performance comparison experiment

Method	Flops (G)	Time (s)	RMSE (°C) ↓	R^2 ↑
ConvLSTM	3.20	3.31	0.5834 ± 0.0434	0.9840 ± 0.0021
PredRNN	14.44	4.46	0.5594 ± 0.0266	0.9856 ± 0.0013
SwinLSTM-B	2.08	4.91	0.5734 ± 0.0292	0.9842 ± 0.0015
SwinLSTM-D	4.28	8.62	0.5604 ± 0.0282	0.9850 ± 0.0013
SA-ConvLSTM	3.79	6.15	0.5159 ± 0.0282	0.9875 ± 0.0014
SimVP	9.92	3.43	0.5191 ± 0.0218	0.9876 ± 0.0010
SWO(ours)	**7.76**	**1.21**	**0.4470 ± 0.0146**	**0.9905 ± 0.0006**



The reconstruction results of different methods at different depths are compared with the measured RMSE results

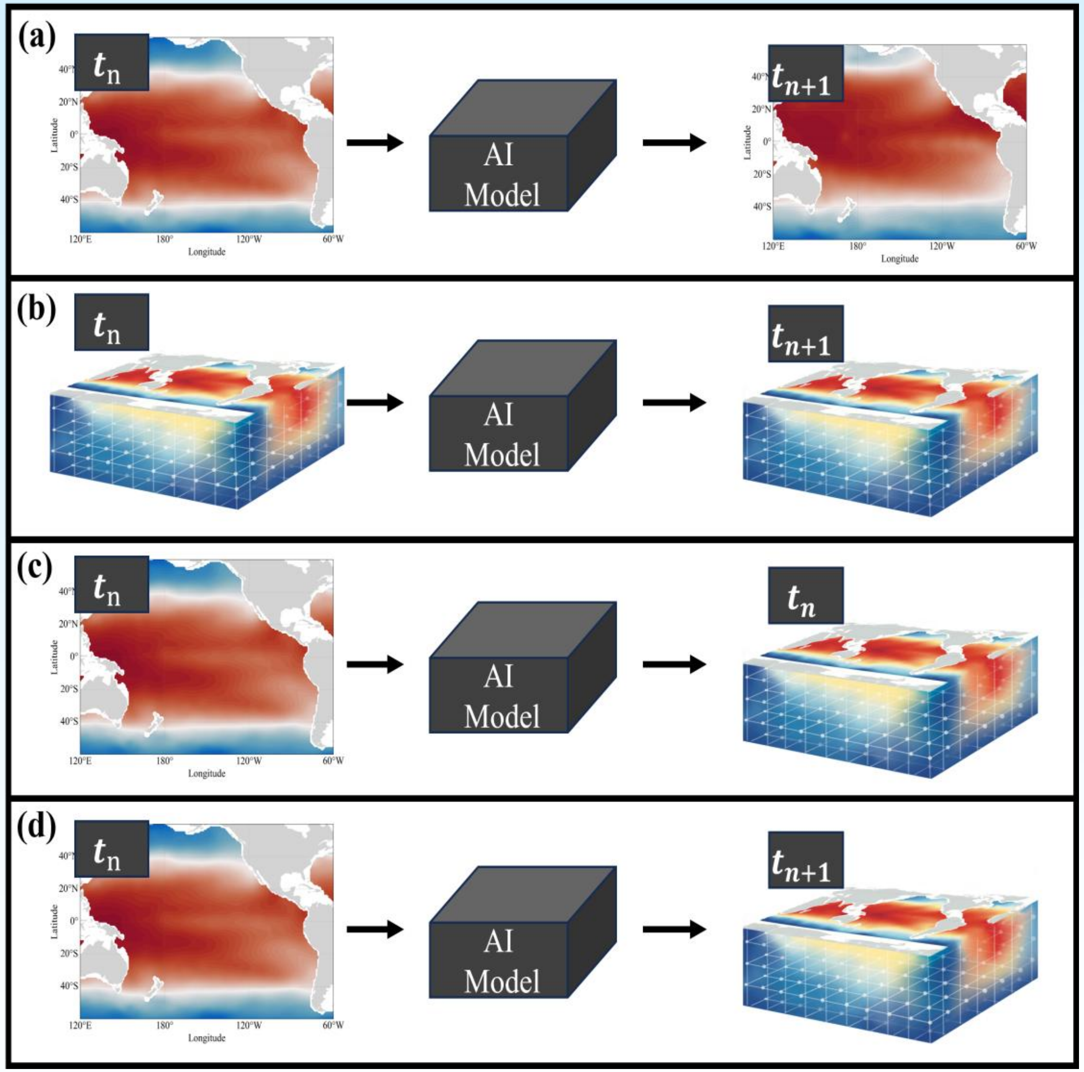




Schematic diagram of the synchronization between satellite remote sensing image changes (SST) and SWO model reconstruction of ocean temperature structure

(a) Compare the RMSE results and measurements of different methods in the Northern Hemisphere for 12 months of 2019. (b) Comparison of seasonal reconstruction results in the Northern Hemisphere.(c) Compare the RMSE results between the 12 months of 2019 and different methods and measurements in the Southern Hemisphere. (d) Comparison of seasonal reconstruction results in the Southern Hemisphere.

construct 3D temperature structure prediction model for the Pacific region based on spatio-temporal AI models



Four forms of Marine environmental structure are studied based on artificial intelligence methods

Predict the future sea surface by using satellite remote sensing data of the sea surface (He et al., 2020; Zhang et al., 2023)

Predict future 3D structures using 3D structures (Sun et al., 2022; Yue et al., 2024)

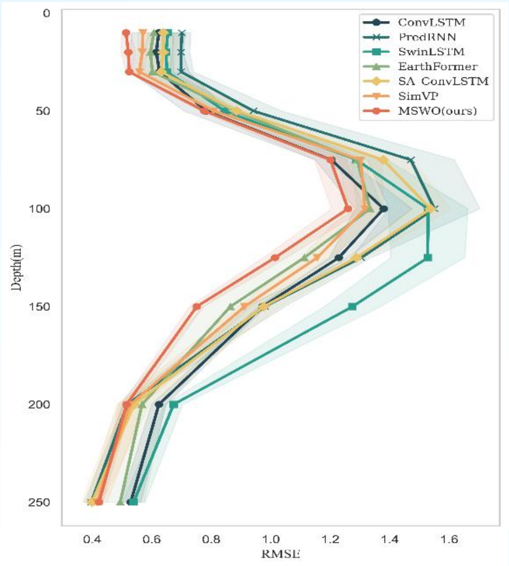
Reconstruct 3D structures using satellite remote sensing data on the sea surface (Liu et al., 2024; Zhang et al., 2024)

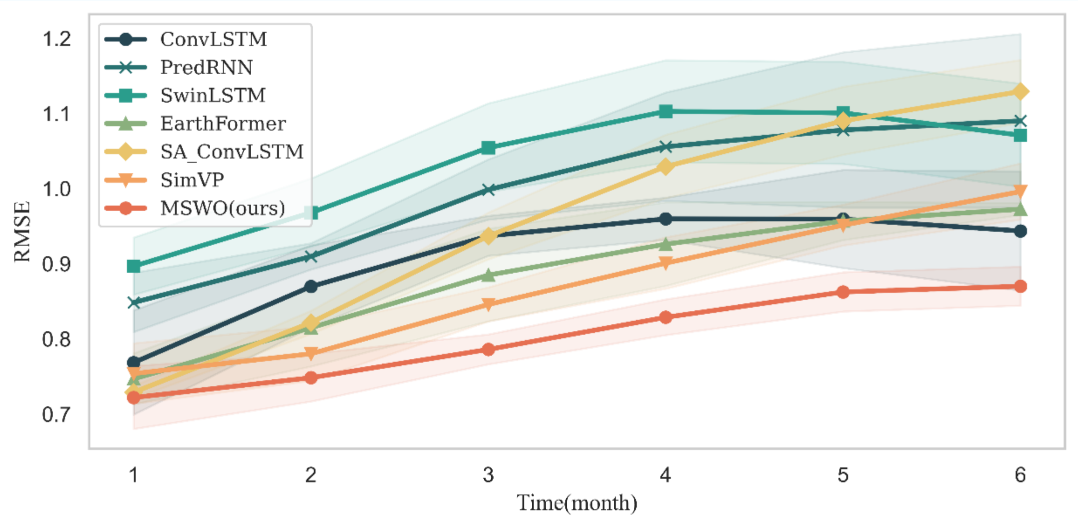
?

Method	Size (H=W)	MSE↓	RMSE↓	MAE↓
ConvLSTM	16	0.829±0.105	0.858±0.060	0.580±0.048
PredRNN	16	1.007±0.143	0.925±0.062	0.634±0.043
SwinLSTM	16	1.074±0.156	0.963±0.064	0.648±0.031
EarthFormer	16/8	0.789±0.091	0.834±0.054	0.596±0.046
SA-ConvLSTM	16	0.935±0.105	0.891±0.054	0.610±0.034
SimVP	16	0.768±0.067	0.813±0.035	0.575±0.023
MSWO (ours)	16/8	0.648±0.047	0.749±0.026	0.516±0.012

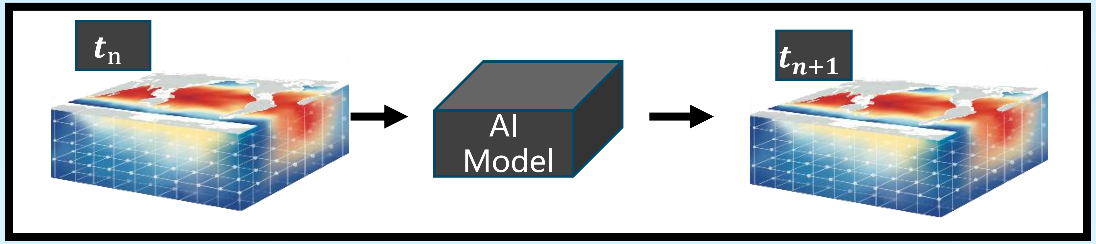
The comprehensive prediction comparison results of the MSWO models and other baseline models on the test dataset

参数名称	参数值
Time resolution	Monthly
Spatial resolution	1°
Dataset interval	2011年-2019年
Source	Argo
The proportion of training, validation, and test datasets	7：1：1
Framework	Pytorch
Loss	MSE
Optimizer	Adam
Batch Size	2
Epoch	1000
Early Stop	50
Learning Rate	0.001
GPU	Nvidia Tesla V100



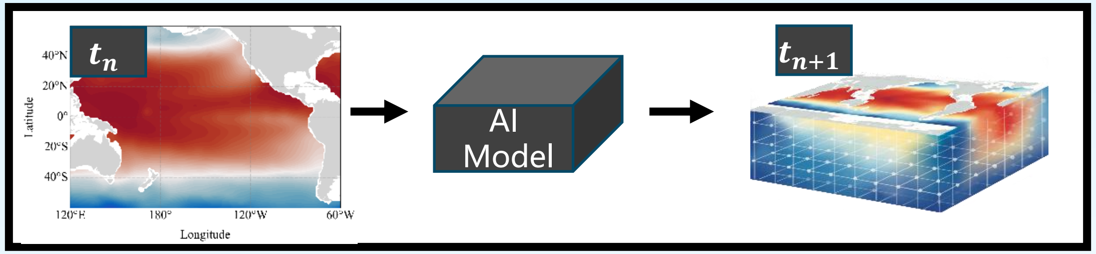


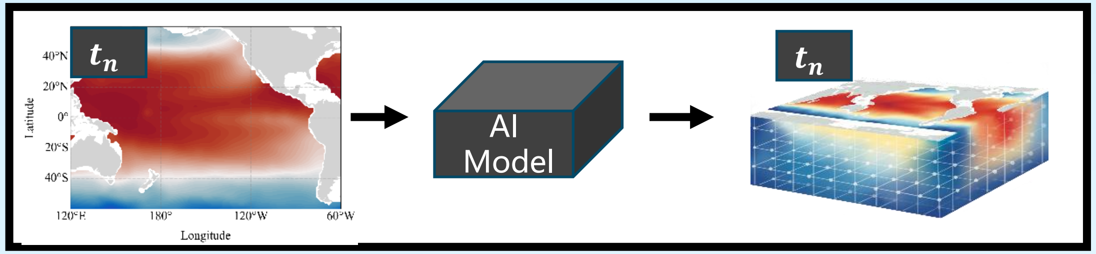
Background Technical route **Discussions**



Predict future 3Dstructures using 3D structures(3D to 3D)
VS
Predict future 3D structures using satellite remote sensing data on the sea surface(2D to 2D)

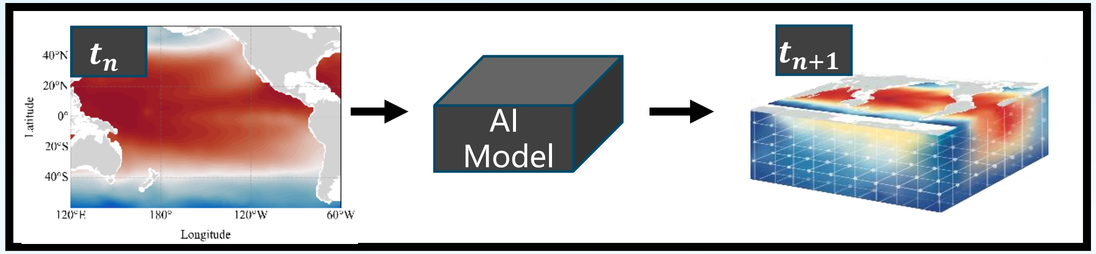
?



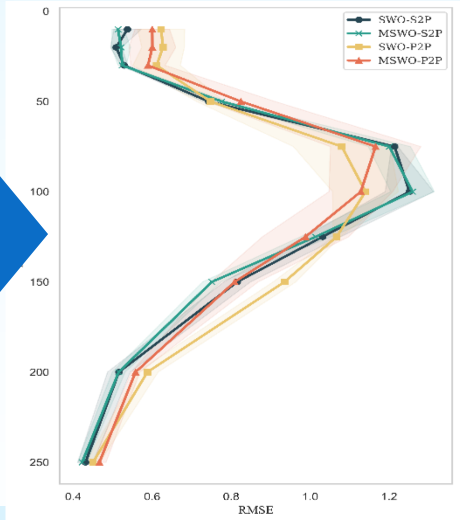


Reconstruct 3D structures using satellite remote sensing data on the sea surface
VS
Predict future 3D structures using satellite remote sensing data on the sea surface

?

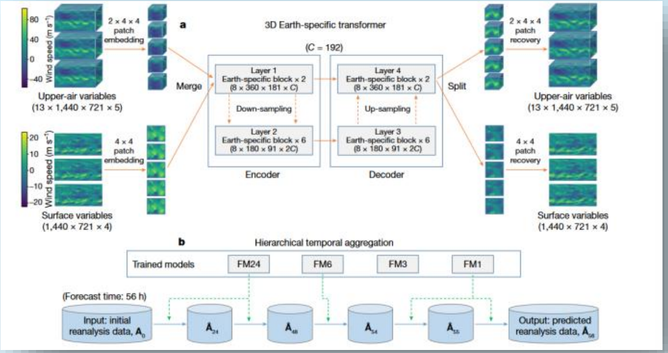


Compare	2Dto3D	3Dto3D
input	Remote sensing	3D numerical field
channel	[B,T,C1,H,W]	[B,T,C2,H,W]
output	3D predictive field	3D predictive field
channel	[B,T,C,H,W]	[B,T,C,H,W]

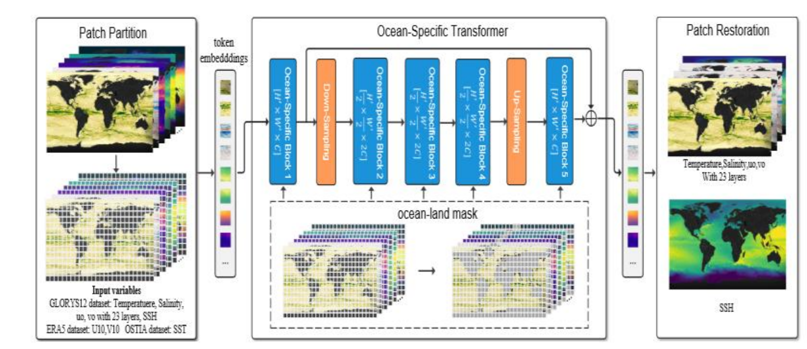


Compare	预报	重建
input	Historical remote sensing	Historical and current remote sensing
channel	[B,T1,C,H,W]	[B,T2,C,H,W]
output	3D field	3D field
channel	[B,T,C,H,W]	[B,T,C,H,W]

Evaluation of high-resolution 3D ocean temperature structure prediction Model based on IV-TT Class4 framework



Pangu: Accurate medium-range global weather forecasting with 3D neural networks



XiHe: A Data-Driven Model for Global Ocean Eddy-Resolving Forecasting

Meteorological and oceanic large models

Features: High resolution, Daily forecast

Comparison object: International mainstream business model systems

Most models are based on the Swin Transformer

Working mode: Input historical 3D field and predict future three-dimensional field (3Dto3D mode)

Advantages: Fast speed and high prediction accuracy

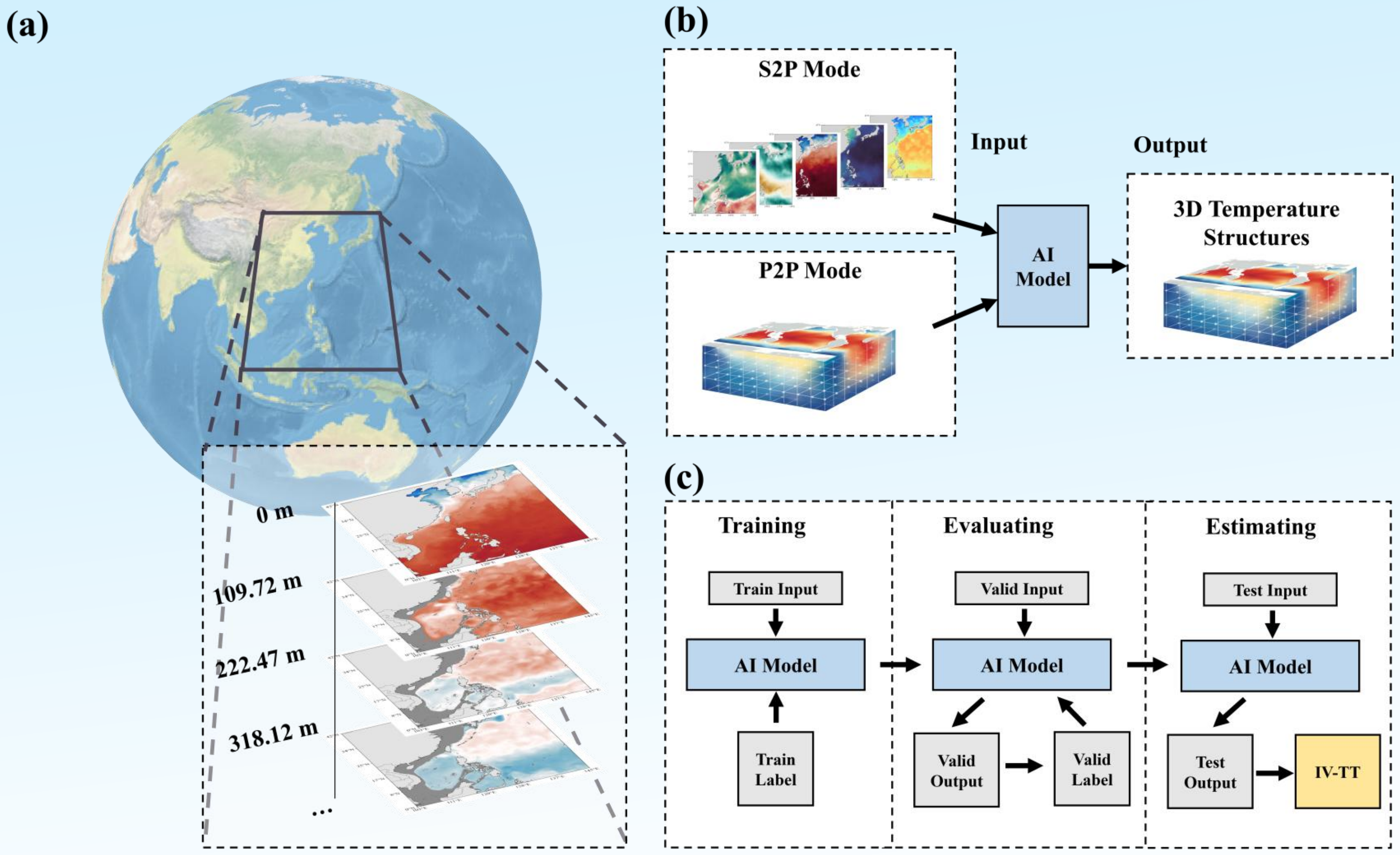
Limitations: High training costs and reliance on model products for historical three-dimensional fields

Can the high-resolution 3D Marine environment be directly predicted using remote sensing images? How effective is it?

Can it be compared with the mainstream international business-oriented Marine forecasting models?

In this work, the input variables of our *XiHe* model include 2 sea surface variables (sea surface temperature and sea surface height), 4 ocean variables with 23 layers (ocean temperature, sea salinity, zonal and meridional component of ocean currents), and 2 atmospheric variables of sea surface (zonal and meridional component of sea surface wind), a total of 96 variables. The sea surface temperature data is from OSTIA and data of other ocean variables comes from the GLORYS12 global reanalysis dataset. The GLORYS12 global reanalysis data has a spatial resolution of 1/12° (4320 × 2041 longitude-latitude grid points) and a temporal resolution of one day. Two atmospheric variables (zonal and meridional components of sea surface wind) are added to help learn ocean dynamics raised by atmosphere forcing, which are from the ERA5 reanalysis dataset from ECMWF. We stack these variables to generate the input data tensor $\mathbf{X}^t \in \mathbb{R}^{W\times H\times C_{\text{in}}}$, where W and H denote the number of grid points in longitude and latitude directions, respectively. C_{in} denotes the number of input variables. With the 1/12° spatial resolution and 96 input variables, we have $W = 4320$, $H = 2041$, and $C_{in} = 96$. Given the input data tensor $\mathbf{X}^t$ corresponding to the current time step, the objective of *XiHe* is to output the K-step-ahead forecasts $\hat{\mathbf{X}}^{t+1:t+K} = \left(\hat{\mathbf{X}}^{t+1},\ldots,\hat{\mathbf{X}}^{t+K}\right)$, $\hat{\mathbf{X}}^{t+\tau} \in \mathbb{R}^{W\times H\times C_{\text{out}}}, \tau = 1,2,\ldots K$. The data dimensions of the output $\hat{X}^{t+\tau}$ are 4320 × 2041 × 94 (2 atmospheric variables are excluded). The forecasting task can be formulated by:

$$\hat{\mathbf{X}}^{t+\tau} = \mathcal{F}\left(\mathbf{X}^t,\theta\right), \tau = 1,2,\ldots,K \qquad (1)$$



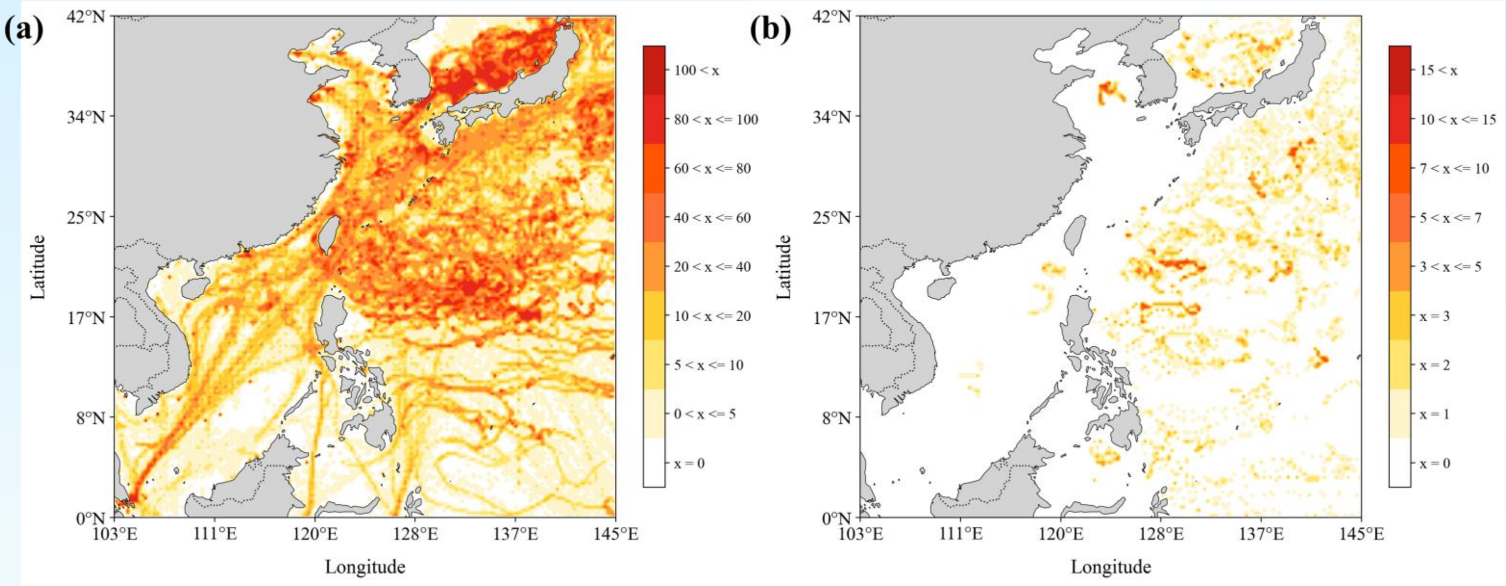
Study area and experimental flowchart. (a) Schematic of vertical temperature profiles in the study area, which encompasses the offshore regions of China and the Western Pacific and ranges from 0° to 42°N and 103°E to 145°E. (b) The distinction between 2Dto3D and 3Dto3D modes lies in the differing input data as input, while both produce 3D temperature structure forecasts as output. (c) The experimental framework consists of three phases—Training, Evaluating, and Estimating—designed to train the model and assess its performance.

参数名称	参数值
Time resolution	Daily
Spatial resolution	1/12°
Dataset interval	2011年-2020年
Source	Glorys12V1
The proportion of training, validation, and test datasets	8：1：1
Framework	Pytorch
Loss	MSE
Optimizer	Adam
Batch Size	2
Epoch	200
Early Stop	30
Learning Rate	0.001
GPU	4*Nvidia 4090

Ocean Predict Intercomparison and Validation Task Team (IV-TT) Class4

The Class 4 standard focuses on the evaluation of system prediction performance by providing observational data such as drifting buoys from Argo buoys and from the USGODAE server for comparative evaluation (Ryan et al., 2015)

The distribution of the measured data of the IV-TT Class4 standard framework in the study area is shown in the following figure. There are 730,423 drifting buoys (left) and 6,275 Argo profiles (right).



System	Institution	Horizontal resolution	OGCM
FOAM	UK Met Office	1/4°	NEMO
GIOPS	Environment Canada	1/4°	NEMO
PSY4	France Mercator Océan	1/12°	NEMO
BLK	Australian Bureau of Meteorology	1/10°	MOM4
MSWO (ours)	IOCAS	1/12°	AI

Experiment 1: Input Parameters of Remote Sensing in 2Dto3D mode

SST, SSS, ADT, UWND and VWND remote sensing data all show positive and active roles in the work of satellite reconstruction of subsurface ocean structures (Meng et al., 2022; Zhang et al., 2024)

Case	Variables	RMSE (°C)↓
Case1	SST, ADT, SSS, UWND, VWND	0.7601
Case2	ADT, SSS, UWND, VWND	0.8534
Case3	SST, SSS, UWND, VWND	0.9551
Case4	SST, ADT, UWND, VWND	0.7868
Case5	SST, ADT, SSS	0.7670

Experiment 2: Performance Comparison of Different Models in 2Dto3D Mode

The prediction results were compared with those of other spatio-temporal prediction models, including ConvLSTM (Shi et al., 2015) and 3D U-Net (Cicek et al., 2016). SimVP (Tan et al., 2023) and SwinLSTM (Tang et al., 2023)

Method	Flops(G)	Time (s)	RMSE (°C)↓
ConvLSTM	24.0606	21.3076	0.8479
3D U-Net	8.5629	22.0464	0.8321
SimVP	5.3883	44.1177	0.8259
SwinLSTM	8.7415	40.3215	0.8033
MSWO (ours)	3.9779	22.4446	0.7601

Experiment 3: Comparison Experiment of Input Time Steps

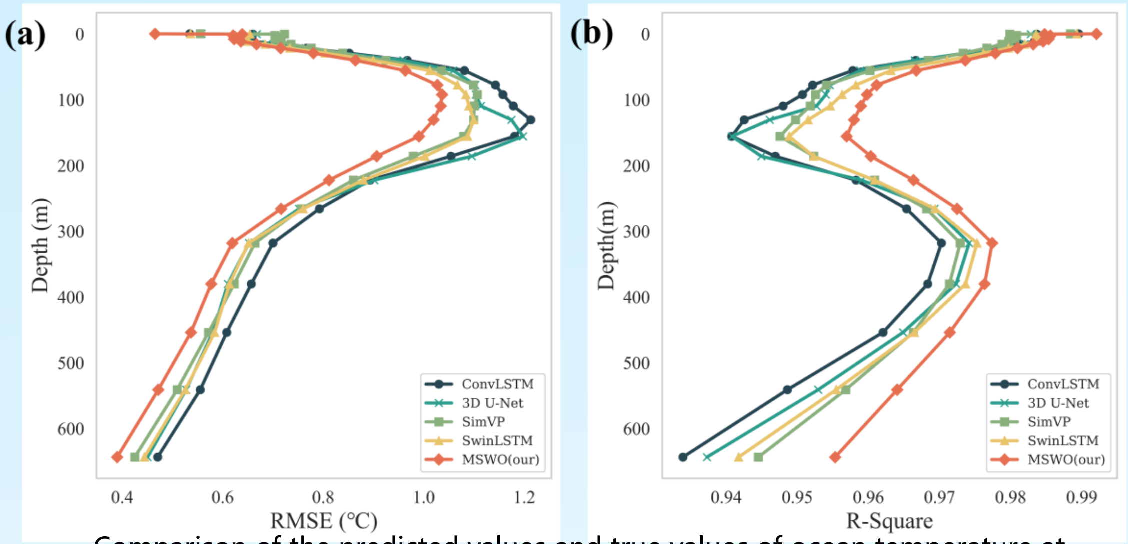
For general time series or spatio-temporal sequence tasks, the introduction of historical data enables the model to learn the changing trends on the time series

Mode	Time Step	RMSE (°C)↓
2Dto3D	1	0.7601
2Dto3D	3	0.7808
2Dto3D	5	0.7677
2Dto3D	7	0.7624
2Dto3D	9	0.7645

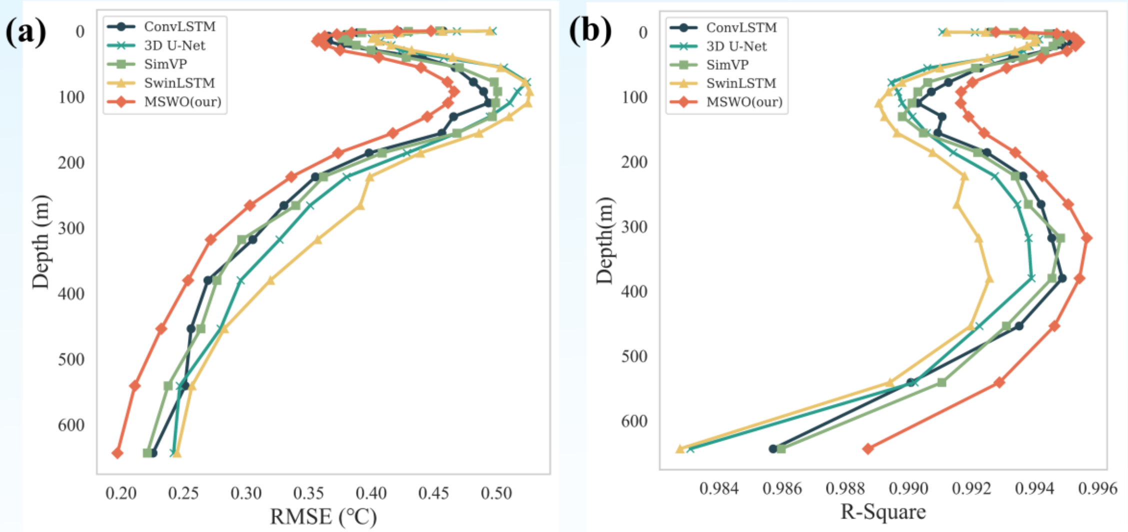
Mode	Time Step	RMSE (°C)↓
3Dto3D	1	0.4232
3Dto3D	3	0.3749
3Dto3D	5	0.3848
3Dto3D	7	0.3724
3Dto3D	9	0.3783

Experiment 4: Performance Comparison of Different Models in 3Dto3D Mode

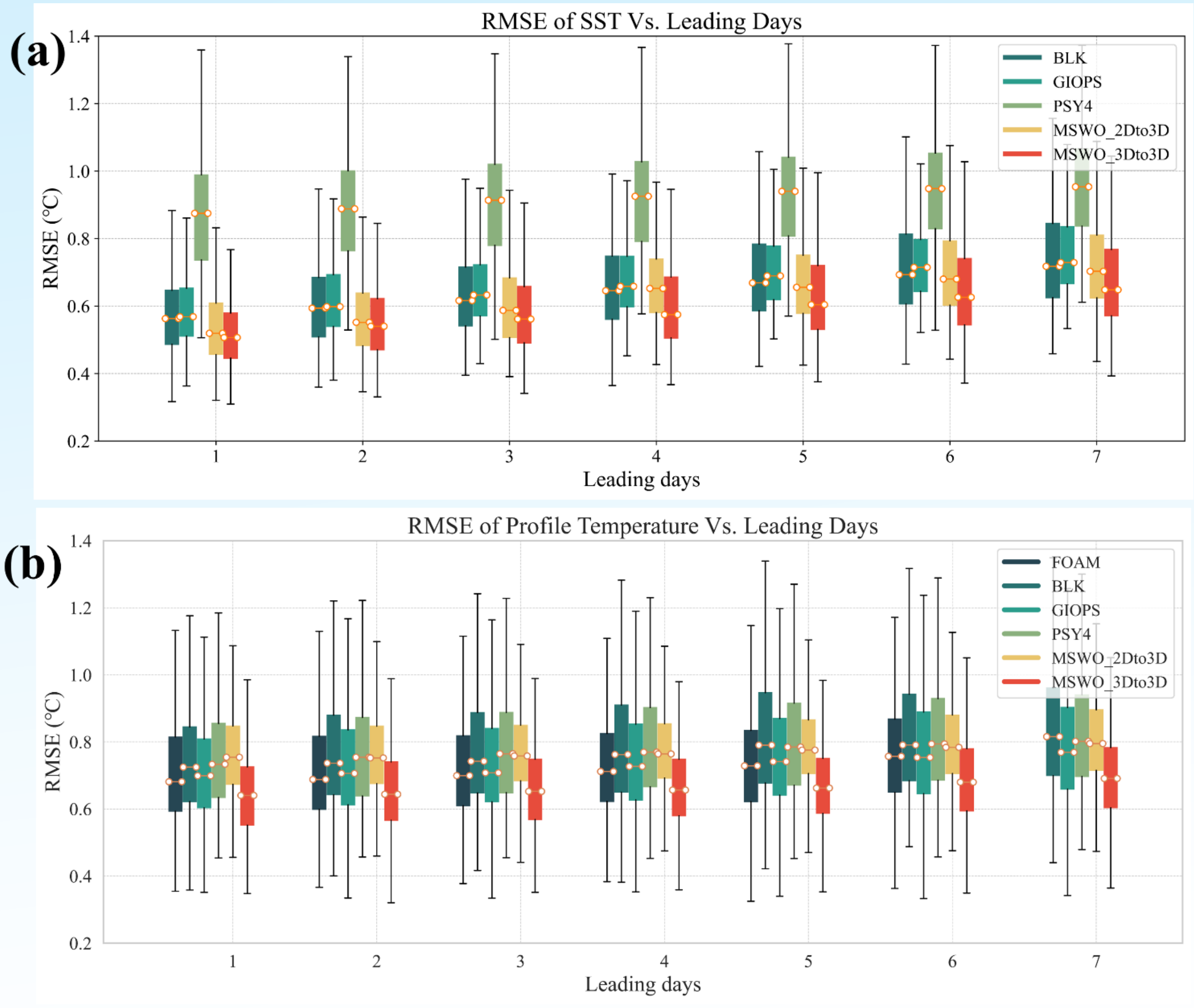
Method	Flops(G)	Time (s)	RMSE (°C)↓
ConvLSTM	31.3597	30.7251	0.3938
3D U-Net	27.0787	21.5309	0.4237
SimVP	15.2906	32.0723	0.4004
SwinLSTM	14.3280	53.0076	0.4298
MSWO (ours)	14.9338	28.0454	0.3749



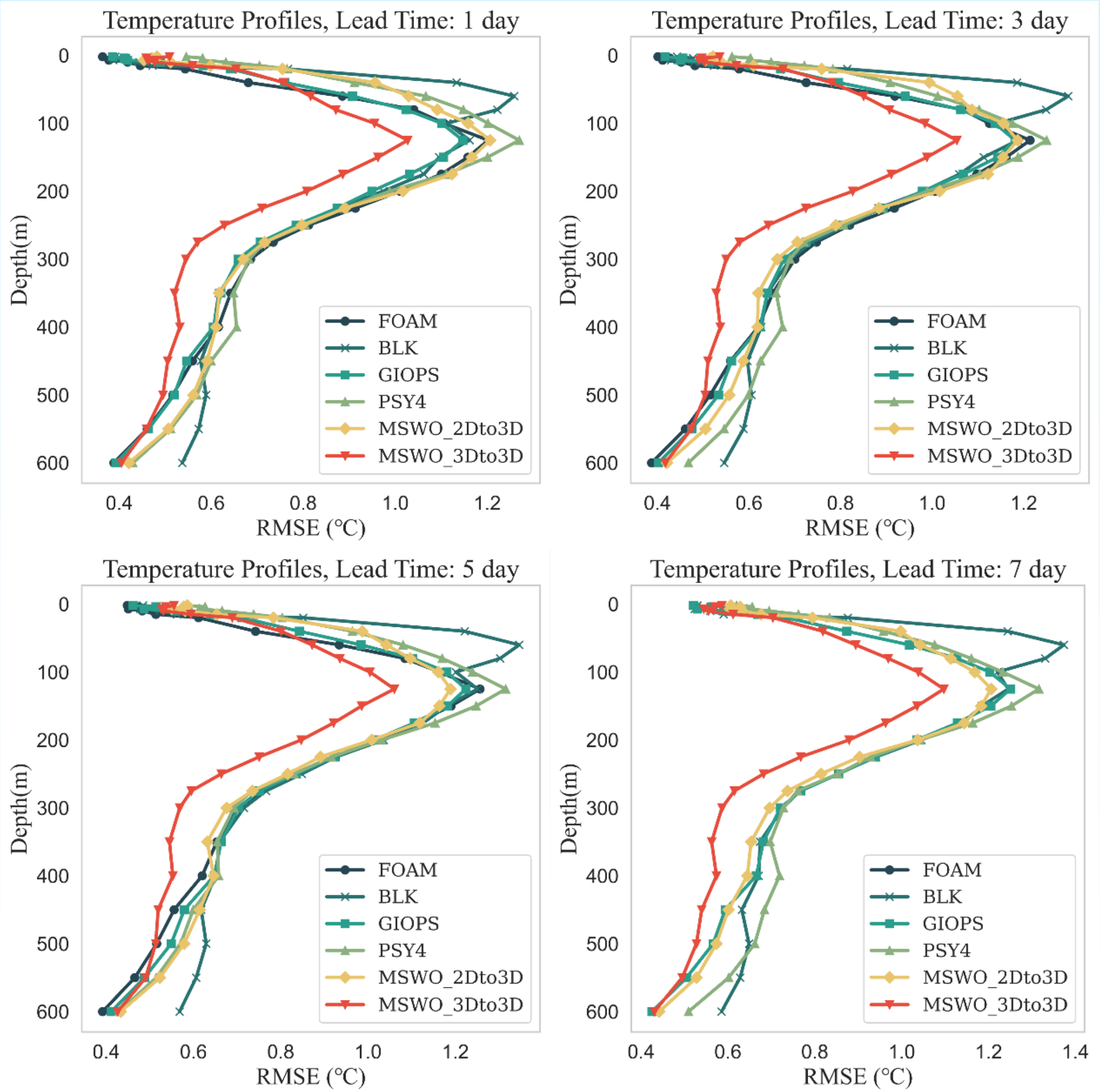
Comparison of the predicted values and true values of ocean temperature at different depths by different methods under the 2Dto3D mode



Comparison of the predicted values and true values of ocean temperature at different depths by different methods under the 3Dto3D mode



Comparison of the MSWO model with different forecasting systems in 2Dto3D mode and 3Dto3D mode.(a) SST, (b) Profile Temperature



The error comparison of the MSWO model and different operational forecasting systems on SST and Profile Temperature from January 1, 2020 to December 31, 2020, with forecasting times of 1 day, 4 days and 7 days respectively. The horizontal axis represents the date variation, and the vertical axis represents the error variation of different model predictions.

1. Construction of Spatio-temporal Sequence AI Model

Based on the previous research on spatio-temporal sequence artificial intelligence models, we have proposed the MSWO model. MSWO combines the sliding window mechanism of Swin Transformer and the spatio-temporal coupling idea proposed in SimVP and efficiently implements spatio-temporal sequence tasks using the window attention mechanism and channel attention mechanism.

2. Construct a 3D temperature structure inversion model for the Pacific region based on spatio-temporal AI model

Reconstructing the 3D structure of the ocean through satellite remote sensing data using AI method is a new approach to achieving transparent ocean observation. The method based on spatio-temporal multi-dimensional attention is an effective means to study this issue. However, with the development of high-resolution satellite images, the computational costs of the spatial and temporal dimensions have become expensive. We used the SWO model and multi-source remote sensing images to reconstruct the subsurface temperature structure of the Central and Southern Pacific. Through experiments, we demonstrated that our model has lower computational overhead and more powerful performance compared to the popular spatio-temporal sequence model.

3. Construction of a 3D temperature structure prediction model for the Pacific region based on spatio-temporal AI model

MSWO introduces a global position encoding method to enrich the spatio-temporal relationship features among data. Through multi-scale residual operations and window attention mechanisms, it extracts spatio-temporal correlations in the input data to achieve prediction and forecast of future subsurface ocean structures. In the comparative experiment, MSWO achieved the best prediction effect, which was attributed to the extraction and efficient utilization of global spatio-temporal information by the attention mechanism. We explored the differences between the 2Dto3D and 3Dto3D models. Surprisingly, the results obtained from these two strategies were approximately the same, which indicates that we can directly predict and forecast future ocean changes through satellite remote sensing means. In addition, we compared the performance of MSWO in the prediction task and the reconstruction task.

4. Evaluation of high-resolution 3D ocean temperature structure prediction Model Based on IV-TT Class4 framework

At present, large-scale oceanographic models based on AI have demonstrated superior forecasting performance. However, these models still rely on the 3D background field generated by numerical simulation as input during forecasting, which brings certain application limitations and relatively high computational costs. We investigated a novel approach that directly uses multi-source remote sensing data to forecast the 3D ocean temperature structure at high spatiotemporal resolution and compared it with traditional ocean forecasting models that depend on 3D marine environmental background fields. The models were also evaluated under the world-leading IV-TT operational forecasting framework. Experimental results show that compared to current mainstream global marine forecasting systems, the MSWO model achieves better forecasting performance in 3Dto3D mode and results comparable to operational forecasting systems in 2Dto3D mode. Compared to other data-driven AI models, MSWO performs best in terms of both training cost and accuracy. This study offers a new perspective for accurate marine environment forecasting, suggesting that satellite remote sensing data could be directly used for high spatiotemporal resolution 3D ocean prediction.

THANK YOU！

Liying Wan