ADVANCING EARTH AND SPACE SCIENCES

JGR Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2026JH001314

Fa Zhang and Hiusuet Kung contributed equally to this work.

Key Points:

- We developed a novel machine learning model to predict long-term and large-scale chlorophyll-a concentrations in the ocean
- The proposed model excels at accommodating the tremendous heterogeneity of chlorophyll-a concentration in the variable ocean
- The partition and model-extracted spatiotemporal characteristics exhibit strong correspondence with the complex dynamics of chlorophyll-a

Correspondence to:
C. Yang and J. Gan,
macyang@ust.hk;
magan@ust.hk

Citation:
Zhang, F., Kung, H., Zhang, F., Wang, Z., Yang, C., & Gan, J. (2026). Taming the heterogeneous dynamics of oceanic chlorophyll-a concentration with a novel deep learning model to improve prediction. *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2026JH001314. https://doi.org/10.1029/2026JH001314

Received 26 FEB 2026
Accepted 28 JUL 2026

Author Contributions:
Conceptualization: Can Yang, Jianping Gan
Data curation: Fa Zhang, Hiusuet Kung
Formal analysis: Fa Zhang
Funding acquisition: Can Yang, Jianping Gan
Investigation: Fa Zhang, Can Yang, Jianping Gan
Methodology: Fa Zhang, Can Yang, Jianping Gan

Taming the Heterogeneous Dynamics of Oceanic Chlorophyll-a Concentration With a Novel Deep Learning Model to Improve Prediction

Fa Zhang[1], Hiusuet Kung[1,2], Fan Zhang[1], Zhiwei Wang[1], Can Yang[1,2], and Jianping Gan[1,2]

[1]Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong, China, [2]Department of Ocean Science, Center for Ocean Research in Hong Kong and Macau, The Hong Kong University of Science and Technology, Hong Kong, China

Abstract Chlorophyll-a (Chl_a) concentration in the ocean is a critical indicator of primary production and plays a pivotal role in the global carbon cycle. Its accurate predictions are essential for reliable projections of future climate change. However, Chl_a is controlled by myriads of interactive physical and biogeochemical processes, posing a substantial challenge for its diagnosis and prognosis through numerical modeling. Machine learning (ML) approaches offer a promising venue to face the challenge, yet existing ML models often struggle in handling Chl_a's significant spatiotemporal variability. We propose a novel machine learning model, Spatiotemporal Dynamics Hunter (STD-Hunter), to tame spatiotemporally heterogeneous dynamics of Chl_a and improve its prediction. STD-Hunter effectively accommodates the heterogeneity in principal dynamics characterized by multiple basis ML models. By integrating these basis models with Chl_a's spatiotemporal characteristics, STD-Hunter forms task-specific predictive models that harvest the underlying dynamics. STD-Hunter optimizes the trade-off between effectively leveraging data and preserving idiosyncratic dynamics. Based on remotely sensing data, we demonstrate the superior performance of STD-Hunter in predicting highly variable Chl_a in an active marginal sea. The extracted characteristics are also well aligned with Chl_a's intrinsic dynamics. By bridging between highly nonuniform observations and their underlying dynamics, STD-Hunter offers an interpretable tool to obtain physically meaningful spatiotemporal characteristics from a data-driven perspective.

Plain Language Summary Chlorophyll-a (Chl_a) concentration in the ocean is a key pigment for photosynthesis in plankton biomass, thus vital in the global carbon cycle. Chl_a is influenced by many different drivers and their interactions, exhibiting great heterogeneity in space and time. This poses significant challenges to its accurate prediction for numerical models and existing machine learning (ML) methods. To handle such heterogeneity, we developed a new machine learning model to predict Chl_a in the ocean over long time scales and large geographic areas. Our new model is called Spatiotemporal Dynamics Hunter (STD-Hunter). The model's core design creates a predictive model by integrating common basis models with Chl_a's spatiotemporal characteristics. Then, there is a training process that jointly reconstructs the spatiotemporal characteristics and the basis models to create a predictive model that performs the single task of providing a value for a specified time and place. We demonstrate how well STD-Hunter performs relative to other preexisting benchmark models using various prediction scenarios. Additionally, we demonstrate the transparency of STD-Hunter and the capability of the mechanism that interprets the data. The model results correspond well to the intrinsic dynamics of the Chl_a distribution in the marginal sea we used as an example.

1. Introduction

Global ocean accounts for 27% removal of the anthropogenic CO_2 emission for the past decade (Friedlingstein et al., 2022), with marine primary production (PP) as the critical component for this carbon sink (Brando et al., 2006). Chlorophyll-a (Chl_a) is a pigment found in phytoplankton biomass, which has been widely used as a proxy for the PP in marine ecosystem (Behrenfeld & Falkowski, 1997; Boyce et al., 2010). Therefore, accurately representing and predicting Chl_a is crucial for better climate forecasting and developing plausible sustainability policies.

To this end, coupled physical-biogeochemical numerical models have been applied on a global scale (Hauck et al., 2020; Lacroix et al., 2021; Long et al., 2021; Wright et al., 2021), at regional scales (Du et al., 2020; Lu

Project administration: Can Yang, Jianping Gan
Software: Fa Zhang
Supervision: Can Yang, Jianping Gan
Validation: Fa Zhang, Hiusuet Kung, Fan Zhang, Zhiwei Wang, Can Yang, Jianping Gan
Visualization: Fa Zhang
Writing – original draft: Fa Zhang
Writing – review & editing: Hiusuet Kung, Fan Zhang, Zhiwei Wang, Can Yang, Jianping Gan

et al., 2020), and at local scales (Gan et al., 2010, 2014; Lu et al., 2022) to diagnose and forecast the physical and biogeochemical evolution of marine ecosystems. However, prediction accuracy of the coupled models is constrained by many uncertainties such as parameterizing the physical sub-grid processes, finding suitable initial conditions (Hofmann et al., 2011), and accounting for intricate biogeochemical interactive processes and parameters (Friedlingstein et al., 2022; Matear, 1995).

With the availability of vast spatiotemporal marine data sets (Esaias et al., 1998; Hauser et al., 2020; Hersbach et al., 2023), machine learning (ML) models, as data-driven approaches, have emerged as an alternative to coupled models for more accurately predicting spatiotemporal trends of Chl_a. ML models have already shown promising performance compared to numerical models in various tasks (Bi et al., 2023; Chen et al., 2024; Lam et al., 2023; Nearing et al., 2024; Zhang et al., 2023). Unlike numerical models, which often rely on simplified assumptions and predefined equations with various approximations, ML models learn directly from data, thereby substantially mitigating the risk of model misspecification. ML models excel at capturing nonlinear relationships between ocean variables and can effectively model complex dynamics. Moreover, their significantly higher computational efficiency enables large-scale predictions of Chl_a.

Machine learning models have proven successful in spatiotemporal time series predictions (Corradini et al., 2024; Torres et al., 2021; Wang et al., 2024; Yang et al., 2024). ML models have also been increasingly developed and applied to predict Chl_a. Park et al. (2015) demonstrated the predictive power of artificial neural network (ANN) and support vector machine (SVM) in Chl_a prediction. Li et al. (2018) and Yajima and Derot (2018) employed the random forest method and examined the feature importance of candidate predictors. Cho et al. (2018) and Zheng et al. (2021) utilized long short-term memory (LSTM) models to capture long-range temporal dependencies in Chl_a concentration prediction. Barzegar et al. (2020) and Liu et al. (2022) integrated convolutional neural network (CNN) with LSTM, forming a hybrid CNN-LSTM model to more effectively extract temporal features. Nevertheless, existing models often overlook the spatiotemporal heterogeneous dynamics inherent in the variables (Shao et al., 2024), so these models struggle in highly variable regions (e.g., Chl_a in the dynamically active ocean). Heterogeneous dynamics lead to variability in the optimal model parameters of ML models in predicting Chl_a in a specific space and at a specific time. The limited observations for each task are insufficient to support the training of complex ML models. The conventional practice of training a single model for all tasks inherently lacks the ability to accommodate the highly heterogeneous dynamics of Chl_a within a single set of parameters.

To overcome the shortcoming presented by heterogeneous dynamics, we propose the novel Spatiotemporal Dynamics Hunter (STD-Hunter) model to tame Chl_a's spatiotemporal heterogeneous dynamics and improve long-term predictions of Chl_a concentration. The core innovation of our model is that STD-Hunter forms a predictive model that integrates commonly shared basis models with specific spatiotemporal characteristics of Chl_a. The basis models and spatiotemporal characteristics are jointly reconstructed using a unified training process. STD-Hunter optimizes the trade-offs between the two conventional approaches. Compared to the approach that trains a separate model for a single task, STD-Hunter employs a low-rank approximation of the parameter tensor to effectively capture the principal dynamics across different tasks and significantly reduces the number of parameters. Compared to the approach that trains a single model for all tasks, STD-Hunter substantially advances the model's ability to accommodate the heterogeneous dynamics of Chl_a into varying model parameters.

In this study, we utilize STD-Hunter to predict the Chl_a concentration in the South China Sea (SCS), the largest marginal sea in the Pacific region. The SCS is characterized by a complex ocean circulation pattern under the influence of the East Asian Monsoon and the exchanges with water masses of the Western Philippine Sea, East China Sea, Sulu Sea, and Java Sea (Gan et al., 2016). In addition to the basin-scale circulation, mesoscale eddies and jets are ubiquitous and display large spatial and temporal variances in both physical and biogeochemical parameters. Inputs from the Pearl River and the Mekong River as well as the adjacent western Pacific Ocean further generate large spatiotemporal gradients in the coastal regions (Ning et al., 2004). The combination of the factors above results in the considerable heterogeneity of Chl_a in this region, posing a huge challenge to accurate prediction.

In the following, Section 2 describes the data and model architecture. Sections 3 and 4 provides the prediction performance and partition provided by the model. Section 5 explores the mechanism and interpretability of the model, followed by the discussion and conclusions in Section 6.

2. Data and Method

2.1. Data

We used chlorophyll-a (Chl_a) concentration and sea surface temperature (SST), which exhibit a strong correlation with Chl-a, as the prediction variables. We obtained Chl_a and SST from the MODIS Aqua projects via https://search.earthdata.nasa.gov/search?q=10.5067/AQUA/MODIS/L3M (Esaias et al., 1998). We used monthly average Level 3 mapped products from January 2003 to December 2020 with a spatial resolution of 4 km. We extracted the data for the deep (>1,000 m) South China Sea (SCS) basin, the largest marginal sea in the western Pacific Ocean that has active biogeochemical dynamics. To save GPU memory for the broad application, we downscaled the extracted data by averaging every 3 × 3 grids, resulting in a spatial resolution of 12 km, and 7806 locations in total. We used the average of historical observations from the OC-CCI data set (Sathyendranath et al., 2019, 2023) to impute the missing data and ignored the interpolated data when evaluating model performance for the test period. Extreme Chl_a events that unfold and dissipate within days or weeks fall outside the primary scope of our long-term, monthly scale prediction setting. To handle irregular extreme values that impede the model training, we winsorized the top and bottom 10% data for the in-sample period. We added a sensitivity analysis on the winsorization threshold in Appendix C. Additionally, we normalized the SST data using the mean and standard deviation of Chl_a data for each location in the in-sample period to align them on the same scale.

The heterogeneous dynamics of Chl_a are shaped using diverse physical and biogeochemical processes. To validate that the characteristics learned by STD-Hunter align with these underlying mechanisms, we utilized a suite of oceanographic and atmospheric proxy variables. We obtained the sea level anomalies and the absolute geostrophic velocity ($\vec{U}_{\text{ocean}}$) for the SCS region from AVISO Ssalto/Duacs altimeter products, produced and distributed by the Copernicus Marine and Environment Monitoring Service (Copernicus Climate Change Service & Climate Data Store, 2018). By calculating the curl of the geostrophic currents, we gained the vorticity. In addition, we obtained the atmospheric related drivers (wind speed, zonal and meridional wind) from the ERA5 products (Hersbach et al., 2020, 2023). We averaged the oceanic and atmospheric drivers monthly and horizontally interpolated these averages to the MODIS grid.

2.2. Model Architecture

STD-Hunter provides an end-to-end ML framework for collaborative long-term and large-scale prediction of Chl_a (Figure 1). This framework forms one model for each predictive task by integrating commonly shared basis models with the specific spatiotemporal characteristics of Chl_a, effectively addressing the challenges posed by Chl_a's heterogeneous dynamics (Figure 1a). STD-Hunter operates in two stages: first, the model jointly recovers the parameter tensor for all tasks, and then, the model independently forms the predictive model for each task.

For each task, S_{ti} predicts Chl_a concentrations at location i and in month t, and the optimal model parameters, $P_{ti} \in R^{M}$, vary greatly, where M is the number of parameters. Together, these P_{ti} form the parameter tensor, $P \in R^{T\times N\times M}$, where T and N are the number of months and locations. Although Chl-a exhibits immense spatiotemporal heterogeneity, its primary variance is driven by a few common underlying dynamics. STD-Hunter leverages this physical intuition by employing a low-rank approximation $\hat{P}$ to estimate P (Figure 1b):

$$\hat{P} = EB + b, E \in R^{T\times N\times K},\ B \in R^{K\times M}, b \in R^{M}.$$

Here, the basis models B capture the core and shared dynamics, while the spatiotemporal characteristics E dictate how the intensity of these shared dynamics varies across space and time. By combining E and B, the rank-K tensor EB accommodates the heterogeneous dynamics of Chl_a. During training, E, B, and the shared bias term b are randomly initialized and jointly optimized. The model imposes smoothness constraints on E to ensure continuous changes through space and time.

STD-Hunter then sends $\hat{P}_{\text{ti}}$ to form the predictive model for task S_{ti} (Figure 1c). During prediction, the 1 × 1 convolutional layer captures the interaction between the input variables (Chl_a and SST data) from the prior 2 years. The transformed features are fed into three temporal convolutional (TC) layers to extract temporal features. Skip connections are integrated to combine features extracted at different levels. The integrated information is then processed by a neural network to form the final prediction. The temporal convolutional network

10.1029/2026JH001314

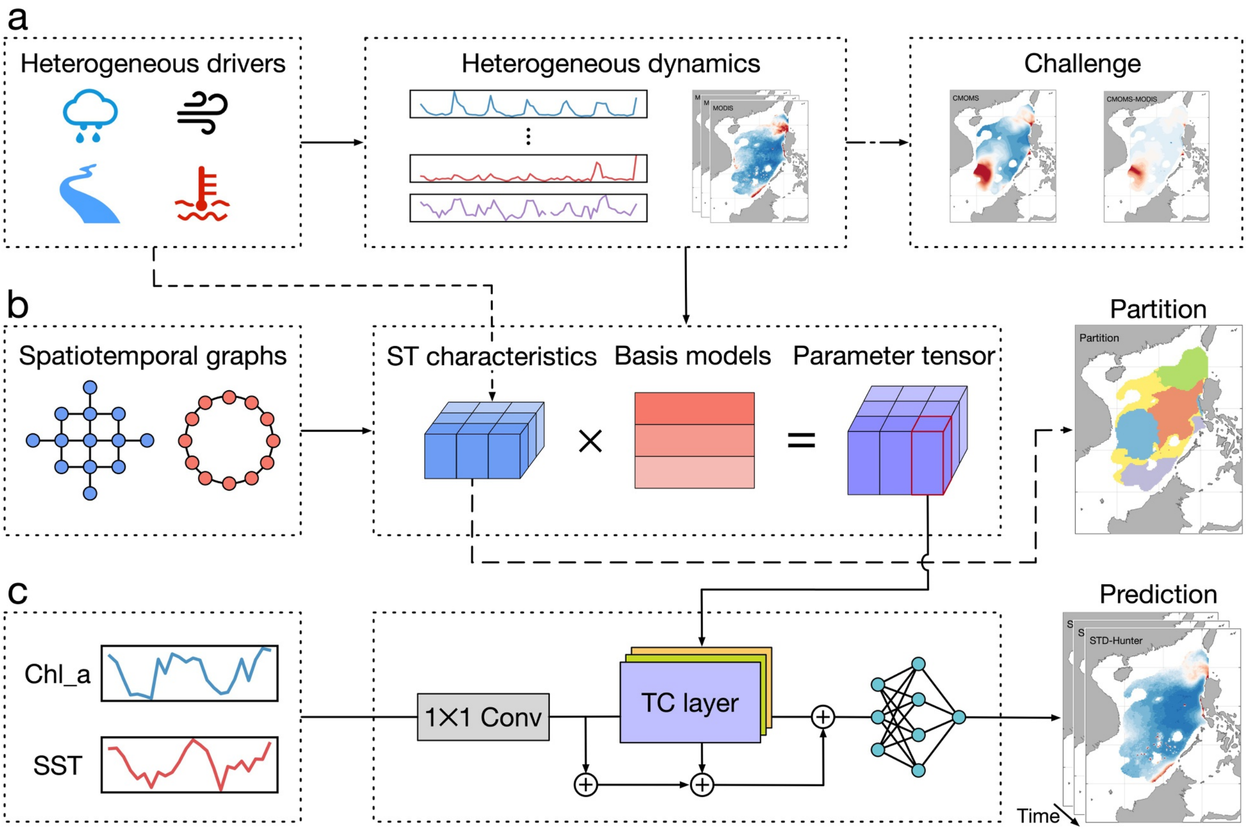

Figure 1. Overall architecture of STD-Hunter. (a) The heterogeneous dynamics of Chl_a pose challenges to accurate prediction. (b) STD-Hunter recovers the parameter tensor through a low-rank approximation. ST characteristics and basis models are randomly initialized, smoothed across space and time, and jointly optimized during training. (c) Predictive model of STD-Hunter. Input data spanning the past 2 years' Chl_a and SST are fed into a 1×1 Conv to capture interactions between the two variables. Three TC layers then extract temporal features across various horizons, with skip connection and residual connection added to preserve intermediate information. Finally, these features are passed into a network to generate the final prediction.

(TCN) in the TC layer is the core component of this predictive model in STD-Hunter, benefited from its efficiency in extracting temporal features of various horizons. The parameters of the TC layers vary across space and time, while others are shared across all tasks. The strategy of adding flexibility solely to core parameters enables efficient use of the increased model capacity. We provide a detailed description of each component below.

2.2.1. Factorization of the Parameter Tensor

To factorize the parameter tensor, the model randomly initializes learnable spatiotemporal characteristics, $E \in R^{T\times N\times K}$. Then, the model passes E to a fully connected layer with a weight matrix, $B \in R^{K\times M}$, and a bias vector, $b \in R^{M}$, to obtain the recovered parameter tensor: $\hat{P} = EB + b$. The low-rank factorization greatly reduces the number of parameters from TNM to $TNK + (K + 1)M$, with E containing TNK parameters, B containing KM parameters and b containing M parameters. Here, all tasks share b, and $E_{\mathrm{ti}}B$ accommodates task S_{ti}'s heterogeneous dynamics. If the heterogeneity across the tasks is insignificant, EB will be small and b will dominate $\hat{P}$, resulting in similar model parameters across the tasks. Otherwise, EB will play more roles to accommodate heterogeneity. This is adaptively adjusted during training.

2.2.2. Smoothness Constraints

The dynamics of Chl_a change smoothly across space and time, leading to a smooth change of P, and, thus, E. Specifically, $E_{t,i}$ is like $E_{t+1,i}$ and $E_{t,i+1}$. The model uses two approaches to impose the smoothness structure on E.

First, the model constructs two adjacency matrices: $A^{\text{time}} \in R^{T\times T}$ and $A^{\text{space}} \in R^{N\times N}$, where $A^{\text{time}}_{t_1,t_2} = 1$ if $|t_1 - t_2| \leq 1$, $A^{\text{time}}_{1,12} = A^{\text{time}}_{12,1} = 1$, and $A^{\text{space}}_{i_1,i_2} = 1$ if $\|c_1 - c_2\|_1 \leq 1$, for c_1 and c_2 the coordinates of space i_1 and i_2. Then, A^{time} and A^{space} are normalized so that each row sums to 1, resulting in $\tilde{A}^{\text{time}}$ and $\tilde{A}^{\text{space}}$. Then, the model performs a graph convolution on E:

$$E = \tilde{A}^{\text{time}} E,\ E_t = \tilde{A}^{\text{space}} E_t, t = 1, \ldots, T.$$

The graph convolution permits each task to share the information in its characteristics with the spatially or temporally adjacent tasks. Additionally, two regularization terms are added to the loss function:

$$\mathcal{L}_1 = \sum_{t=1}^{T} \sum_{i,j=1}^{N} \|E_{ti} - E_{tj}\|^2 \tilde{A}^{\text{space}}_{i,j}, \mathcal{L}_2 = \sum_{t=1}^{T} \|E_t - E_{t+1}\|^2, (E_{T+1} = E_1).$$

The regularizations penalize differences in the characteristics of spatially or temporally adjacent tasks, reducing the flexibility of E.

By applying these approaches, the spatiotemporal characteristics exhibit the following pattern: characteristics of spatially or temporally adjacent tasks are alike to prevent overfitting idiosyncratic noise, and characteristics of spatially and temporally nonadjacent tasks can differ significantly, providing the flexibility to accommodate heterogeneity.

2.2.3. Temporal Convolutional Layer

The Temporal Convolutional Network (TCN) uses 1D convolution to effectively capture temporal patterns and dependencies in time series and has been demonstrated to compete well with recurrent architectures such as LSTMs and GRUs (Bai et al., 2018; Lea et al., 2016, 2017; Van Den Oord et al., 2016; Yan et al., 2020). TCN offers several advantages over RNN-based architectures, including parallel computing, lower memory costs, and stable gradients (Bai et al., 2018). We based the TC layer of STD-Hunter on a TCN variant (Wu et al., 2020) that incorporates gating mechanisms and employs multiple kernel sizes. Furthermore, our model performs circular padding before temporal convolution to minimize information loss. Additionally, the model independently applies transformations to each channel within a skip connection to reduce the number of parameters.

The input features are the past 2 years' Chl_a and SST. To better capture recurrent intra-annual seasonal dynamics, these inputs are reformulated into a matrix of shape $[4, T]$, where $T = 12$. We then stack these feature matrices across all $N = 7806$ locations, yielding a tensor of shape $[4, N, T]$. Utilizing a mini-batch training strategy with a batch size of $b = 12$, the final input data fed into the model has the shape $[b, 4, N, T]$. First, the input data is passed through a 1×1 convolutional layer to obtain C_0 of shape $[b, 4d, N, T]$, where $4d = 16$ is the number of channels. Subsequently, C_0 is fed into three TC layers, whose architecture illustrated in Figure 2. All the parameters within TC layers, including those of convolutional filters, gating mechanisms, and residual connections, are generated by passing task-specific characteristics into fully connected layers. In Figure 2a, each 1D convolution employs d 1D convolutional kernels to extract diverse temporal features. The gating mechanism is employed to pass the important information and suppress irrelevant information. Within the TC layer (Figure 2b), C_0 is passed separately into four 1D convolutions with kernel sizes of 2, 3, 6, and 7, resulting in D_0 of shape $[b, 4d, N, T]$, concatenated by four outputs of shape $[b, d, N, T]$. Subsequently, D_0 is processed in two ways. First, the features of each channel are aggregated through a linear transformation, obtaining S_0 of shape $[b, 4d, N, 1]$. Then, S_0 is added to the skip connection to retain information at this level. Second, D_0 is passed through a 1×1 convolutional layer to mix temporal features across different horizons, resulting in C_1 of shape $[b, 4d, N, T]$. C_1 serves as the input for the next TC layer. STD-Hunter stacks three TC layers to enable 1D convolutions in different TC layers to progressively combine and extract temporal features of various horizons (Figure 2c).

2.3. Settings

We used the past 2 years' Chl_a and SST data from the SCS as model input, and the future Chl_a was our prediction target. The training period was from 2003 to 2014. The validation period was from 2015 to 2016, and

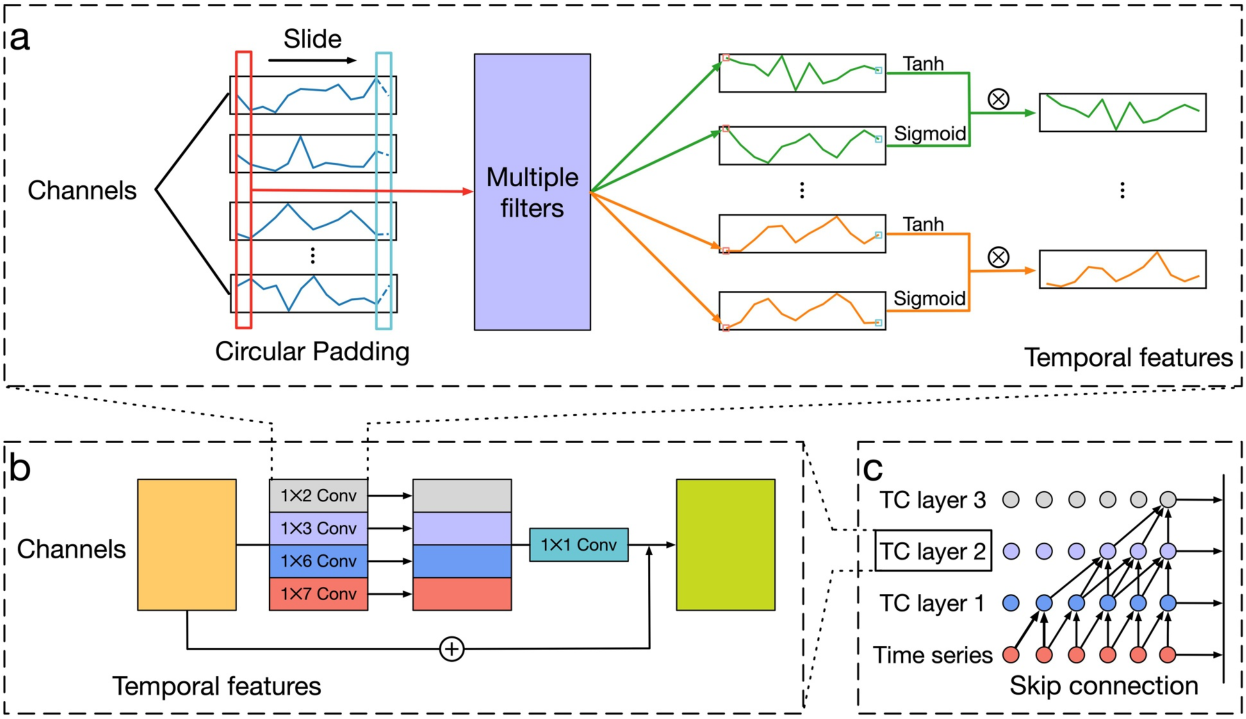

Figure 2. Design of the TC Layer. (a) Inner operations of the 1D convolution, where multiple filters extract temporal features and the gating mechanism controls the information flow. (b) The architecture of the TC layer, featuring several 1D convolutions with different kernel sizes and residual connection for training stability. (c) The stacking of several TC layers extracts temporal features of different horizons.

the test period was from 2017 to 2020. Specifically, given a reference month T, the model takes a continuous 24-month lookback window [T-23, T] of Chl_a and SST to predict the Chl_a target at a future month $T + \tau$. Here τ denotes the prediction horizon, set to $\tau = 12$ for main results and $\tau \in \{24, 36, 48, 60\}$ for longer horizons. The Mean Absolute Error (MAE) between the model prediction and the ground truth was the loss function, and we used the Adam optimizer with a learning rate of 0.001 to minimize the loss and update the model parameters. We used the validation set to select the optimal training epoch and prevent overfitting. For each epoch, we recorded the model's performance on the validation set and its parameters and selected the model with the best validation performance for predictions in the test period.

We used Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Correlation (CORR) as metrics to assess the predictive accuracy of our prediction, $\hat{y}$, when compared to the ground truth, y:

$$\text{MAE}: \frac{\sum_{t=1}^{T} |\hat{y}_t - y_t|}{T}, \ \text{RMSE}: \sqrt{\frac{\sum_{t=1}^{T} (\hat{y}_t - y_t)^2}{T}}.$$

$$\text{MAPE}: \frac{1}{T}\sum_{t=1}^{T} \left|\frac{\hat{y}_t - y_t}{y_t}\right|, \ \text{CORR}: \frac{\sum_{t=1}^{T} (y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})}{\sqrt{\sum_{t=1}^{T} (y_t - \bar{y})^2 \sum_{t=1}^{T} (\hat{y}_t - \bar{\hat{y}})^2}}.$$

The smaller the better for MAE, RMSE, and MAPE, and the bigger the better for CORR. For each location, we computed these metrics between the model prediction $\hat{y}$ and the ground truth y in the test period. Then, we took the average for each metric across all locations to obtain the overall model performance.

Lastly, we compared STD-Hunter with several competitive ML benchmark models: XGBoost (Chen & Guestrin, 2016), LSTM (Hochreiter & Schmidhuber, 1997), CNN-LSTM (Barzegar et al., 2020), Graph WaveNet (Wu et al., 2019), and MTGNN (Wu et al., 2020) (see Appendix A for details).

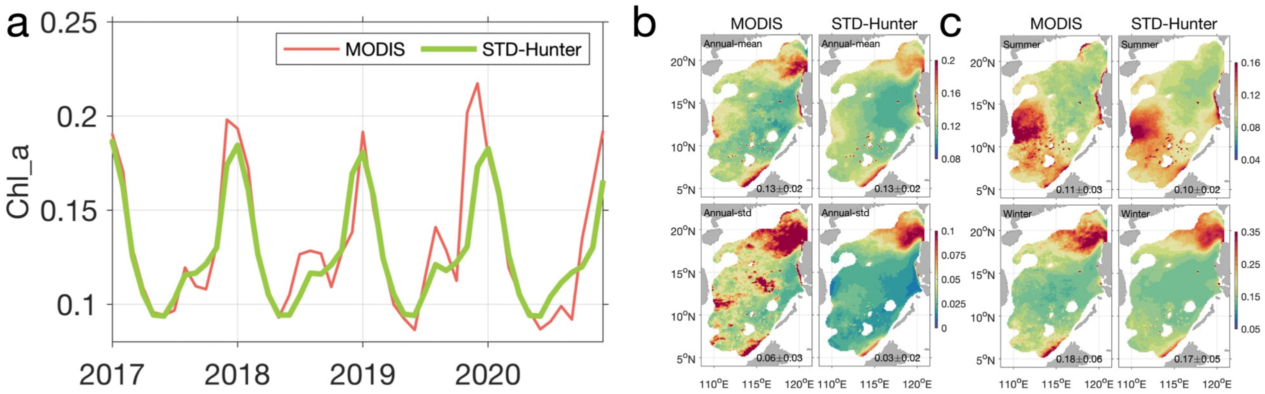

Figure 3. (a) Time series of domain-averaged observed and predicted Chl_a ($\mu g^{-1}L$) from 2017 to 2020 in the SCS. (b) Horizontal distribution of annual mean and standard deviation of observed Chl_a from MODIS and predicted Chl_a from STD-Hunter. (c) Horizontal distribution of seasonal mean of observed Chl_a from MODIS and predicted Chl_a from STD-Hunter.

3. Results

3.1. Prediction of Chl_a in the SCS

We assessed how well STD-Hunter performed on a 1-year prediction of Chl_a concentration in the SCS and compared the results with results from the benchmark models (Figures 3 and 4). The time series of observed domain-averaged Chl_a (Figure 3a) showed that Chl_a concentration in the SCS basin displayed a strong seasonal cycle, peaking in winter and reaching the minimum during spring, with a second peak during summer. The annual mean distribution of Chl_a (Figure 3b) revealed several phytoplankton active zones in the SCS: near the Luzon Strait, west of the Philippine, east off the Vietnam coast, and in the southeastern flank of the deep basin (Palacz et al., 2011). The maxima near the Luzon Strait contributed the most to the winter peaks, while the one off Vietnam was the main contributor to the small peaks during the summer season (Figure 3c). All these spatio-temporal features were well aligned with the underlying coupled physical-biogeochemical processes in the SCS (Lu et al., 2020).

STD-Hunter captured the seasonal variation of the domain mean Chl_a concentration and the annual averaged Chl_a horizontal distribution well (Figures 3a and 3b). Although the peak values during December were smaller in the prediction, the averaged value over the winter months during the test periods (December–February) was very close to the MODIS data (0.18 ± 0.06 $\mu g^{-1}L$ for MODIS, and 0.17 ± 0.05 $\mu g^{-1}L$ for STD-Hunter, Figure 3c). Similarly, STD-Hunter reproduced the horizontal pattern of the temporal mean during the summer months (June–August), even though the second peak showed much larger variations among different years (Figure 3a). The spatial distributions of correlation, MAE, MAPE and RMSE between the STD-Hunter predicted values and the observed values are given in the left column of Figure 4c. Generally, temporal correlations were larger than 0.5, except for the regions near the Vietnam coast and its extended middle basin where Chl_a was largely affected by variable nutrient discharge from the Mekong River. The best performance took place in the northern part of the basin, where the Chl_a's periodicity was more pronounced. Large MAE and RMSE values corresponded to the phytoplankton active regions (west to Luzon, east off Vietnam, and the southeast flank of the basin). Meanwhile, MAPE, which includes the relative bias compared to the Chl_a value, was more pronounced in the Vietnam coast and less significant in the other two regions.

To evaluate STD-Hunter against benchmark models, we computed the domain mean improvements. For each location, we calculate the percentage reduction in error metrics (MAE, MAPE and RMSE) and the percentage increase for CORR. Averaging these improvements across all locations yields the domain mean improvements. STD-Hunter improved the performance of temporal correlation by 7%–14% (Figure 4b), which was more distinct in the central and southern parts of the SCS (Figure 4c), a region where Chl_a's periodicity was less pronounced. With regard to the error, the MAE, MAPE, and RMSE of STD-Hunter were better than the error metrics of the five benchmark models (Figure 4c), with the MAPE differing the most (3%–8%). Improvements mostly occurred near the Luzon Strait and in the southeastern part of the SCS basin (Figure 4c). The STD-Hunter ranked highest with regard to predicting temporal and spatial features (Figure 4a). In general, STD-Hunter predicted better than

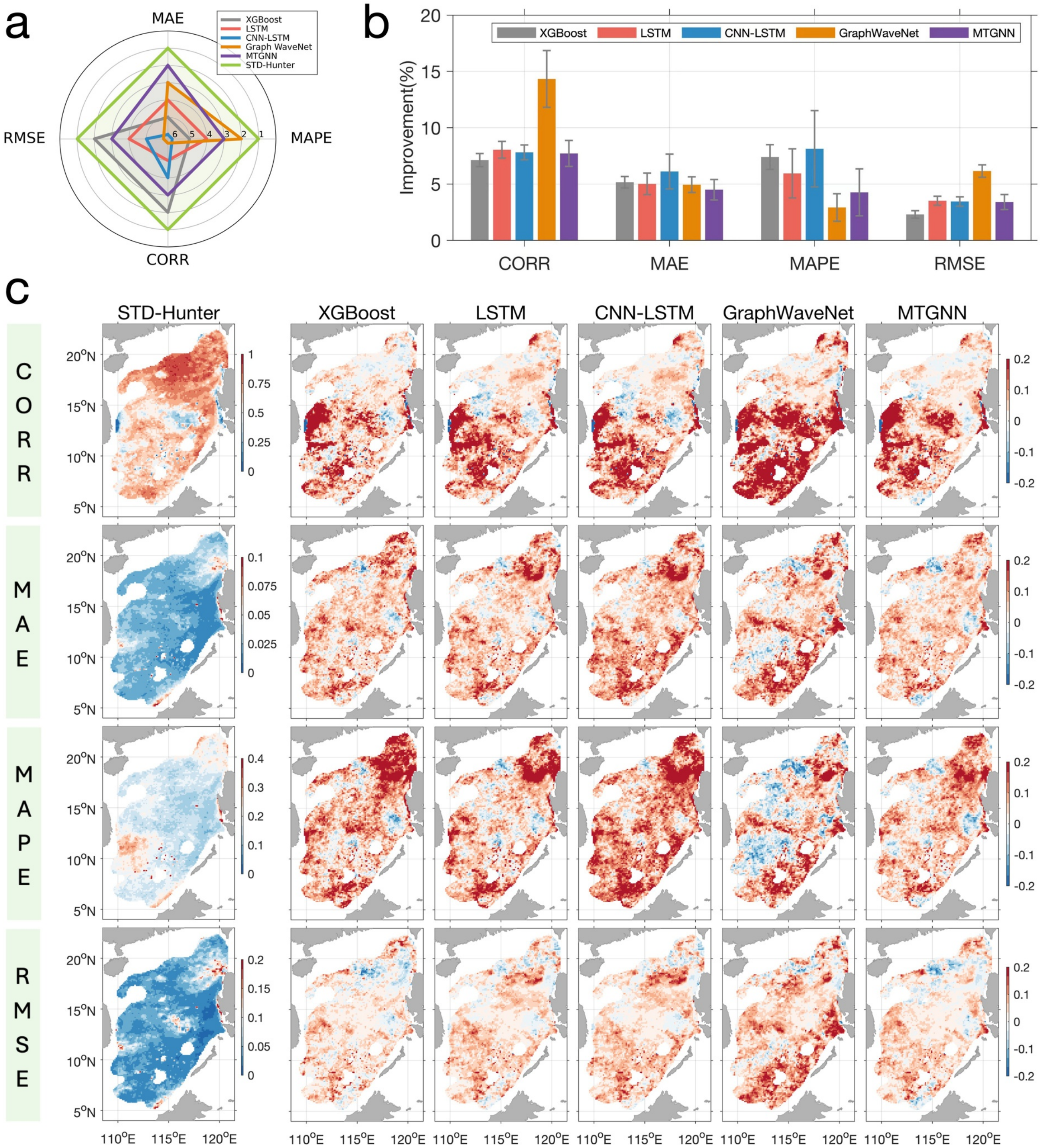

Figure 4. (a) Overall ranking graph of 6 models using metrics CORR, MAE, MAPE, and RMSE. (b) Domain mean improvement of STD-Hunter compared to benchmark models for CORR, MAE, MAPE, and RMSE. The mean and standard deviation of 10 independent runs are reported (c) Horizontal distribution of STD-Hunter performance (left column) and the improvement of STD-Hunter relative to XGBoost (2nd column), LSTM (3rd column), CNN-LSTM (4th column), GraphWaveNet (5th column), and MTGNN (6th column). The top row is the temporal correlation, and the 2nd to 4th row are MAE, MAPE, and RMSE, respectively.

the benchmark models, especially in regions where heterogeneity in space and/or time was more pronounced, such as Chl_a in the SCS margins where the coupled physical-biogeochemical forcing was more variable.

3.2. Further Experimental Results

We further evaluated the performance of the STD-Hunter for longer prediction horizons. We maintained the same model inputs, and the identical division of training, validation and testing periods. We ranged the prediction target

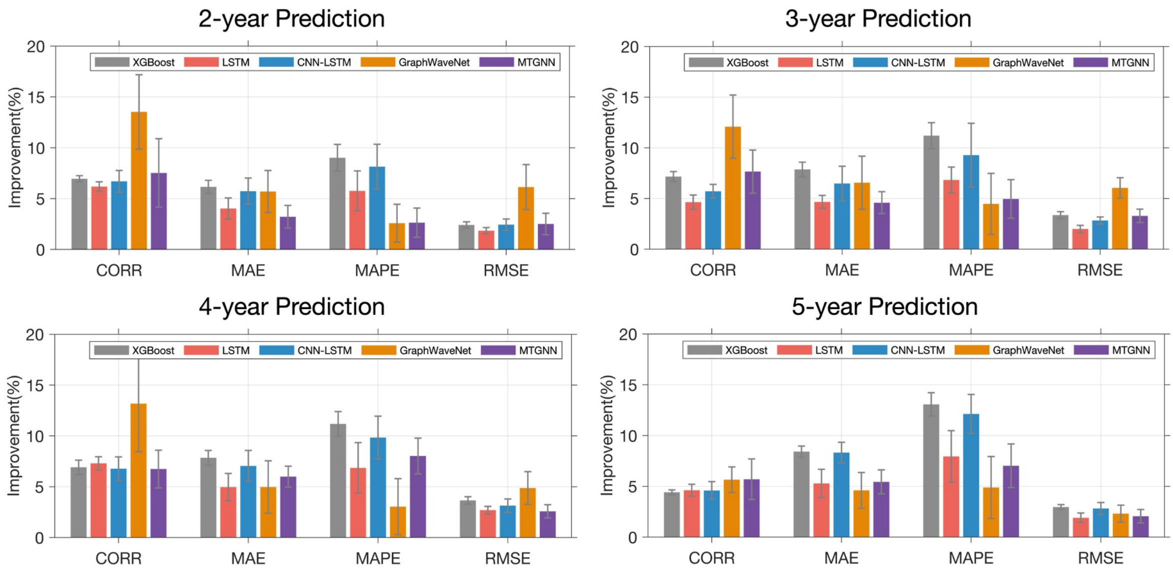

Figure 5. For longer prediction horizons, from the next 2 to 5 years (2 and 3 years in the first row, 4 and 5 years in the second row), the domain mean improvement of STD-Hunter to other benchmarks for CORR, MAE, MAPE, and RMSE in the testing period from 2017 to 2020. The mean and standard deviation of 10 independent runs are reported.

from Chl_a in the next 1 year to Chl_a in the next 2 to 5 years. The results (Figure 5) indicate that STD-Hunter continues to outperform other benchmarks in terms of various metrics. The magnitude of these improvements was similar to the 1-year prediction results, indicating STD-Hunter's ability to capture long-term trends and fluctuations in Chl_a.

We also conducted ablation studies to verify the effectiveness of each component of our model by testing several scenarios:

1. w/o graph layer: without graph convolutions on characteristics E.
2. w/o regularization: without two smoothness regularization terms.
3. w/o SE: without spatial characteristics, thus the shape of characteristics E becomes $T \times 1 \times K$.
4. w/o TE: without temporal characteristics, thus the shape of characteristics E becomes $1 \times N \times K$.
5. concatenate SE: instead of using characteristics E to generate parameters of TCN, simply concatenate the learnable spatial characteristics of shape $1 \times N \times K$ with the input features, and all tasks share one TCN model.
6. concatenate TE: simply concatenate the learnable temporal characteristics of shape $T \times 1 \times K$ with the input features.

Results (Figure 6a) indicated that both the spatial and temporal characteristics of STD-Hunter played a pivotal role, supporting the argument that heterogeneity exists across both space and time. Moreover, imposing smoothness constraints on the characteristics E proved beneficial to model performance. In contrast, the naïve approach of directly concatenating spatial or temporal characteristics with input features failed to enhance model efficacy, underscoring the necessity of factorizing the parameter tensor P.

Furthermore, we conducted sensitivity tests to demonstrate the robustness of model performance to changes in hyperparameters. For each hyperparameter, we made slight adjustments around the default setting and reported the results from 10 independent runs. Specifically, we evaluated the following configurations: dropout (0.3, 0.7), number of channels (8, 32), number of layers (2, 4), weight decay (1e−7, 1e−5), K (2, 8) and regularization (1e−6, 1e−4). The results (Figure 6b) illustrate that STD-Hunter is robust to variations in hyperparameters.

4. Partition of the SCS

The biological productivity in the SCS is highly variable, subject to the controls of spatiotemporally heterogeneous physical and biogeochemical forcing. Identifying the variability is crucial to advancing scientific

10.1029/2026JH001314

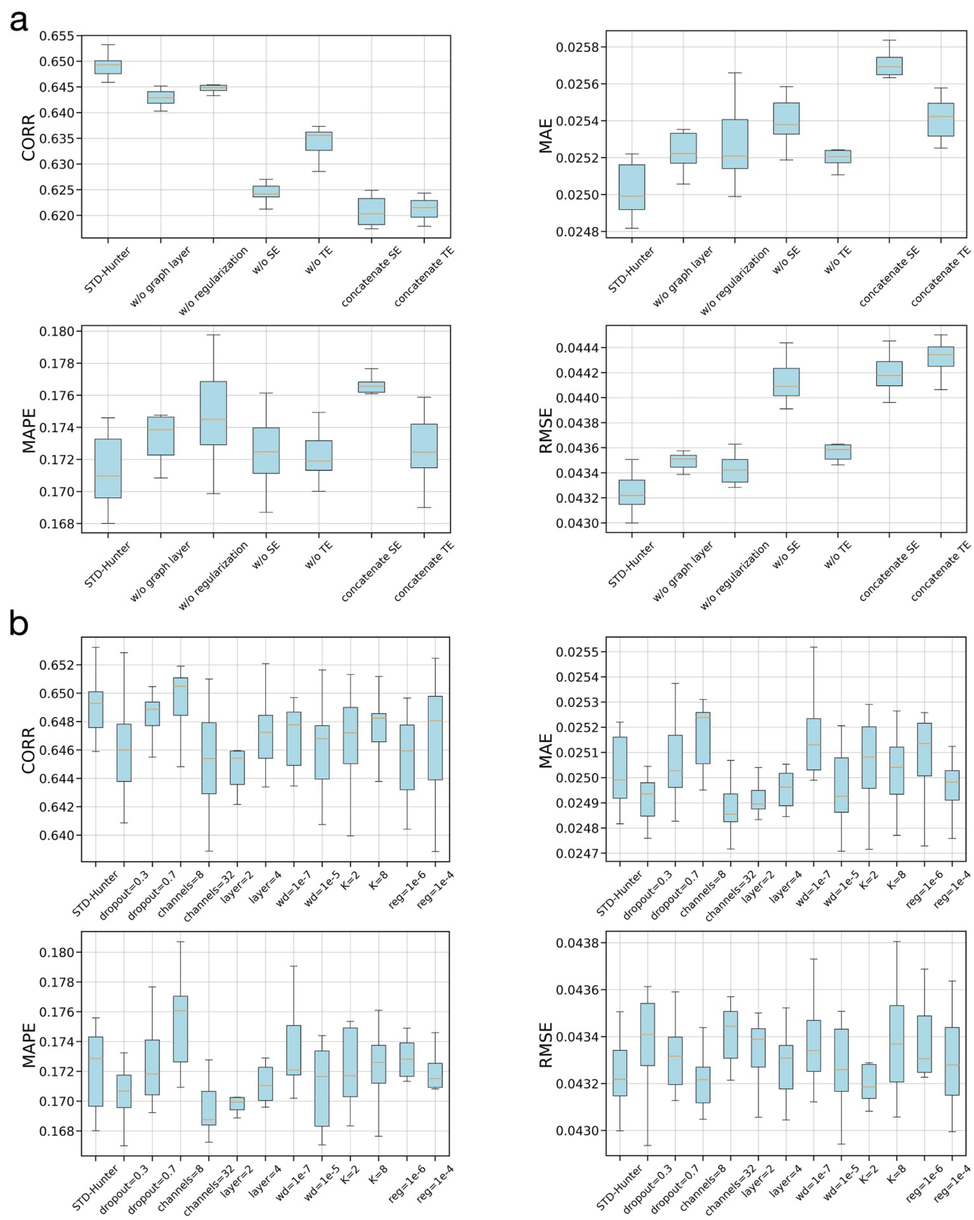

Figure 6. (a) Ablation studies of each component of STD-Hunter. The results of 10 independent runs are reported. (b) Sensitivity tests on the model performance with respect to various hyperparameters. Each hyperparameter was adjusted individually. The results of 10 independent runs are reported.

understanding of ocean ecosystem and underlying mechanisms. STD-Hunter accommodated the heterogeneous dynamics of Chl_a within the varying model parameters, which were determined by the spatiotemporal characteristics E. We flattened E along the temporal dimension to obtain the spatial characteristics $E^{\text{space}} \in R^{N \times TK}$ to identify the variability, which otherwise would not have been feasible. This approach provided us with an invaluable diagnosis of the spatial distribution of Chl_a in the SCS.

We divided the SCS into five regions by applying the K-means clustering algorithm to E^{space} (Figure 7a). STD-Hunter offers a fresh perspective on the partition of the SCS based on Chl_a's intrinsic dynamics rather than direct observations (Xu et al., 2020).

To demonstrate these dynamics, we calculated how various drivers correlated with Chl_a for each of the five regions (Figure 7b). In general, SST had the highest correlation with the Chl_a concentration, followed by wind,

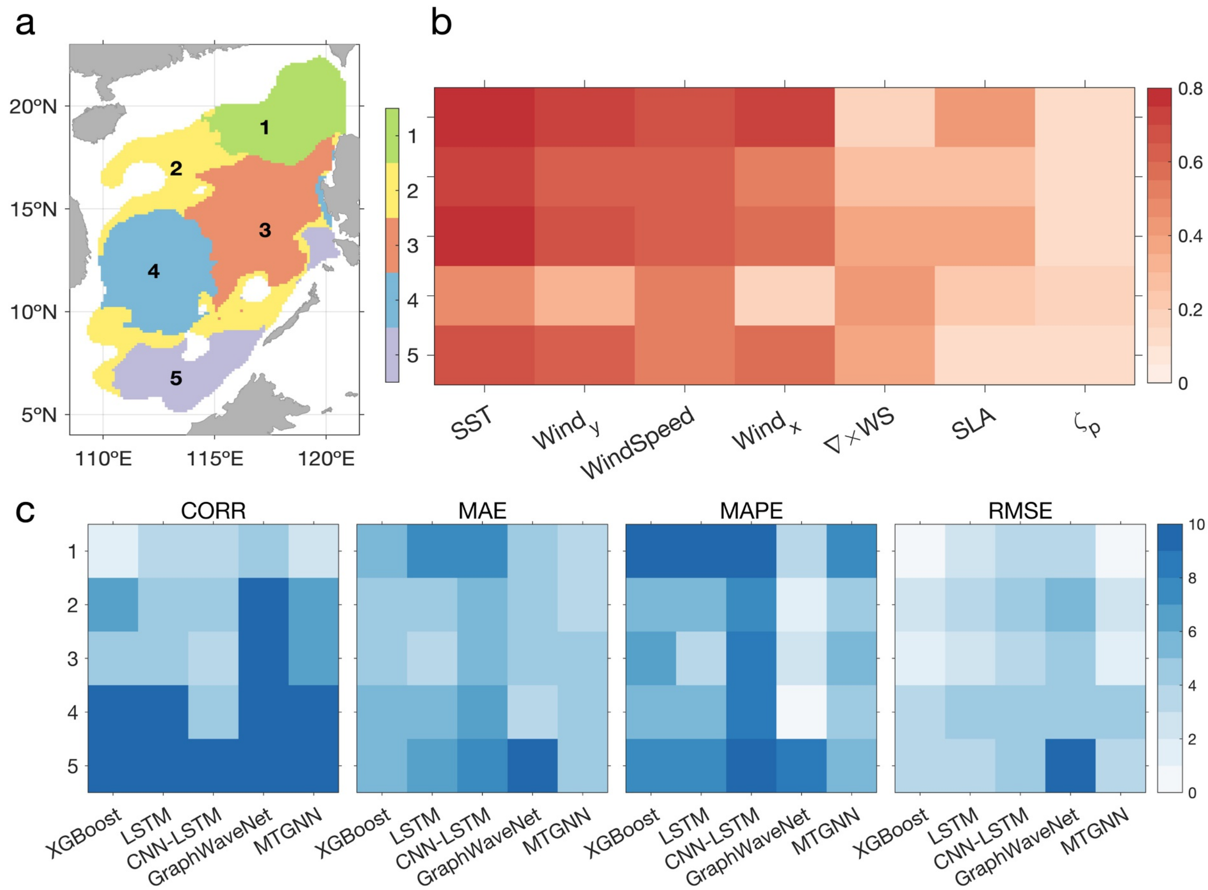

Figure 7. (a) Partition map. (b) Correlation of various factors with Chl_a observations for the 5 regions. (c) Improvement of STD-Hunter to benchmark models with regard to CORR, MAE, MAPE, and RMSE for the different regions.

sea level anomaly (SLA), and ocean current related factors. In the eastern and southern parts of the basin (Regions 1, 3, and 5), the wind speed related components ($\mathrm{Wind}_{\mathrm{speed}}$: mean wind speed at 10 m; $\mathrm{Wind}_x/\mathrm{Wind}_y$: wind speed along zonal/meridional direction) correlated best with Chl_a, which was related to the monsoonal wind-generated nutrient pumping because of mixing from the depths as discussed in Palacz et al. (2011) and Lu and Gan (2023). The curl of the wind ($\nabla \times \mathrm{WS}$), a driver related to Ekman pumping that modifies the Chl_a concentration by transporting higher nutrient water upward, was more significant in Regions 3, 4, and 5, where the summer monsoon dominates. At the same time, low-nutrient water intrusion from the western Pacific Ocean through the Luzon Strait affected the eastern part of the basin (Regions 1 and 3). The intrusion adjusted the circulation pattern via pressure gradient (SLA) and modified the Chl_a with the oligotrophic Pacific water, resulting in a higher correlation between the eastern SCS and Pacific waters.

We also examined the vorticity ($\zeta = \nabla \times \vec{U}_{\mathrm{ocean}}$ where ζ_p represents $\zeta > 0$) of the ocean currents ($\vec{U}_{\mathrm{ocean}}$), a variable associated with the upward transport of water mass and nutrients. As expected, the largest impact from vorticity occurred in regions east of Vietnam's coast (Region 4), where the vorticity related to jet separation is greatest (Gan et al., 2022). Among all the regions, Region 4 was the most difficult region to predict Chl_a due to its complex dynamics and low periodicity. In this region, Chl_a was greatly influenced by variable discharges of high nutrient waters from the Mekong River and ensuing offshore transport. Both wind and its related ocean currents had a low correlation with Chl_a. The influences of multiple factors interacted nonlinearly and reduced the periodicity in this region, raising the challenge to better prediction. However, STD-Hunter identified the correlated characteristics in this region and showed the second-best improvement in correlation (Figure 7c).

10.1029/2026JH001314

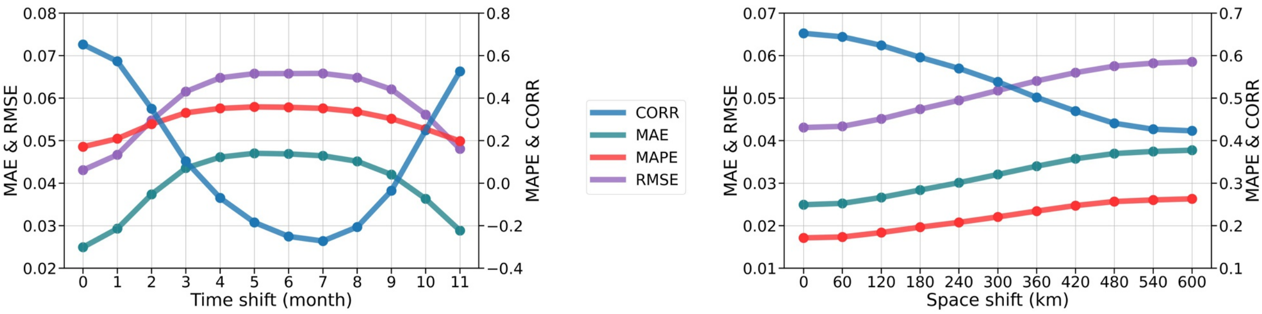

Figure 8. Performance after shifting $\hat{P}$ along spatial or temporal dimensions.

5. Mechanism and Interpretation From Spatiotemporal Characteristics

5.1. STD-Hunter Accommodates Heterogeneous Dynamics

We mainly attribute STD-Hunter's superior performance to the model's ability to substantially mitigate the limitations of conventional practices that rely on a single set of parameters for all tasks.

First, STD-Hunter effectively recovers the optimal parameters for each task by leveraging a low-rank approximation of the parameter tensor P. We examined how well STD-Hunter performed when $\hat{P}$ was shifted along spatial or temporal dimensions (Figure 8). Performance deteriorated remarkably with larger shifts, highlighting the significant variability in optimal model parameters across space and time. The performance declined the most around shifts of 6–7 in time, corresponding well to the Chl_a's seasonally varying signals. In addition, with increasing space shift, all performance criteria worsen. The deterioration was most obvious with shifts of ∼100–500 km in space, indicating the large spatial variability due to oceanic meso-scale (∼100 km) structures in the SCS. This underscores the crucial role of applying nonuniform parameters in highly heterogeneous environment predictions, in which the conventional approach is inherently suboptimal.

Second, STD-Hunter fosters a balanced training scheme that enables tasks to simultaneously achieve optimal performance. We compared the training schemes of STD-Hunter and One-set-of-parameters (OSP). OSP is a variant of STD-Hunter that employs a single set of parameters by nullifying spatiotemporal characteristics. For each training epoch, we performed an empirical orthogonal function (EOF) analysis of the fitted Chl_a, which decomposed the data's heterogeneous features into a linear combination of multiple modes. We focused on the first two EOF modes, which explained the most (71.6%) of the variance. We used the relative error between the EOF modes of Chl_a and the fitted Chl_a to quantitatively compare the training scheme and the testing behavior for each approach:

$$\text{relative\_error}_1 = \frac{\left\| \hat{y}^{\text{mode1}} - y^{\text{mode1}} \right\|_1}{\left\| y^{\text{mode1}} \right\|_1}, \text{relative\_error}_2 = \frac{\left\| \hat{y}^{\text{mode2}} - y^{\text{mode2}} \right\|_1}{\left\| y^{\text{mode2}} \right\|_1}.$$

In general, the optimal relative errors for both EOF modes were higher for OSP, highlighting the distortion of heterogeneous dynamics within a single set of parameters. During the training period, STD-Hunter simultaneously accommodates the two distinct modes by leveraging its basis models and spatiotemporal characteristics, while OSP optimizes the second mode much more slowly than the first because the capacity is not sufficient (Figure 9a). The different training schemes lead to distinct testing behaviors (Figure 9b). For STD-Hunter, the two modes achieved optimal testing performance at similar epochs, reflecting the model's ability to efficiently capture distinct dynamics. However, for OSP, the inefficient and protracted learning process resulted in dispersed epochs for the two modes to achieve optimal performance.

For both approaches, we decompose the model's prediction error into three components (see illustration in Appendix B):

$$\begin{aligned} \text{error} &= \text{error}_{\text{optimal}} + (\text{error}_{\text{suboptimal}} - \text{error}_{\text{optimal}}) + (\text{error} - \text{error}_{\text{suboptimal}}) \\ &= \text{error}_{\text{optimal}} + \text{dispersion} + \text{mismatch}. \end{aligned}$$

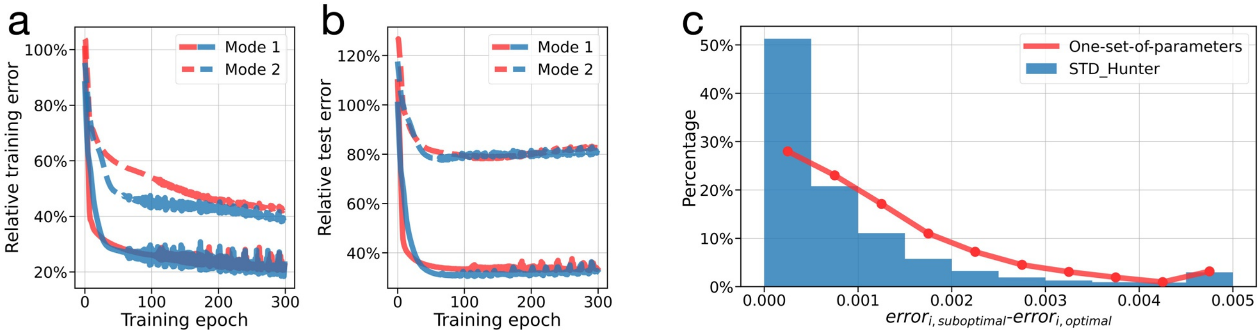

Figure 9. (a) Relative error between the EOF modes of Chl_a and the fitted Chl_a during the training period. (b) Relative error during the test period. (c) The histogram of $\text{error}_{i,\text{suboptimal}} - \text{error}_{i,\text{optimal}}$ across locations. For location i, $\text{error}_{i,\text{suboptimal}}$ was evaluated at $\text{epoch}_{\text{optimal}}$, and $\text{error}_{i,\text{optimal}}$ was evaluated at $\text{epoch}_{i,\text{optimal}}$.

Here, $\text{error}_{\text{optimal}}$ represents the potential optimal error, where each location achieves its optimal performance at $\text{epoch}_{i,\text{optimal}}$. STD-Hunter achieves a lower $\text{error}_{\text{optimal}}$, due to its ability to recover superior model parameters across tasks. $\text{error}_{\text{suboptimal}}$ represents the minimal overall prediction error, achieved at $\text{epoch}_{\text{optimal}}$. The term dispersion quantifies the suboptimality arising from the inability to simultaneously achieve optimal errors across all locations. By optimizing distinct dynamics simultaneously, the STD-Hunter exhibits a more concentrated distribution of optimal performance across locations (Figure 9c). The component mismatch comes from the discrepancy between $\text{epoch}_{\text{optimal}}$ and the epoch selected during the validation period. The reduction in overall error achieved by STD-Hunter stems from reductions in the three components, with $\text{error}_{\text{optimal}}$ contributing 36.1%, dispersion 53.4%, and mismatch 10.5%.

5.2. STD-Hunter Captures Principal Dynamics

STD-Hunter effectively accommodates the first K principal dynamics of Chl_a across space and time in different spaces spanned by the basis models. Its objective of minimizing prediction error leads to a low-rank approximation $\hat{P} = EB + b$ of the parameter tensor P, the proxy of Chl_a's dynamics. For $K = 1$, STD-Hunter captures the first principal dynamics that explain the majority of the variance in P along the basis model B, while other dynamics may be distortedly accommodated within the remaining model capacity. As K increases, additional principal dynamics unexplained by the previously identified dynamics is then gathered and accommodated into the new space. By accommodating distinct principal dynamics in different spaces spanned by the basis models, STD-Hunter enables their effective and accurate representation.

To investigate the principal dynamics, we reformulated $\hat{P}$ as follows:

$$\hat{P} = EB + b = E^{\text{PCA}}A^{-1}B + b_E B + b = E^{\text{PCA}}B^{\text{PCA}} + b^{\text{PCA}}.$$

Here $E^{\text{PCA}} = (E - b_E)A$ represents the PCA transformation applied to E, whose variation reflects the heterogeneous dynamics of Chl_a across space and time. For $K = 4$, the first two components of E^{PCA} explained the majority of variance in E (93.0% in total), indicating that the first two principal dynamics dominate Chl_a's heterogeneous dynamics. B^{PCA} represents the basis models of the K principal dynamics. E^{PCA} characterizes the strength of these principal dynamics. Unveiling their spatial distribution and temporal evolution offers a window into the complex and varying dynamics of Chl_a in the SCS. The characteristics of the first principal dynamics (Figure 10) indicated highly diversified monthly changes in intensity among various active regions. In the Luzon strait active zone, the winter bloom associated with local eddies (Gan et al., 2022) dominated and the large value lasted from November to May in the coming year. Meanwhile, activity in the east off Vietnam coast became notable from May and intensified during the summer when southeasterly monsoon, high Mekong River discharge and strong coastal upwelling prevailed.

We demonstrated the validity of the spatiotemporal characteristics by revealing their intrinsic connections with the spatial patterns of Chl_a's dominant EOF modes. The parameter tensor serves as a proxy of the dynamics of

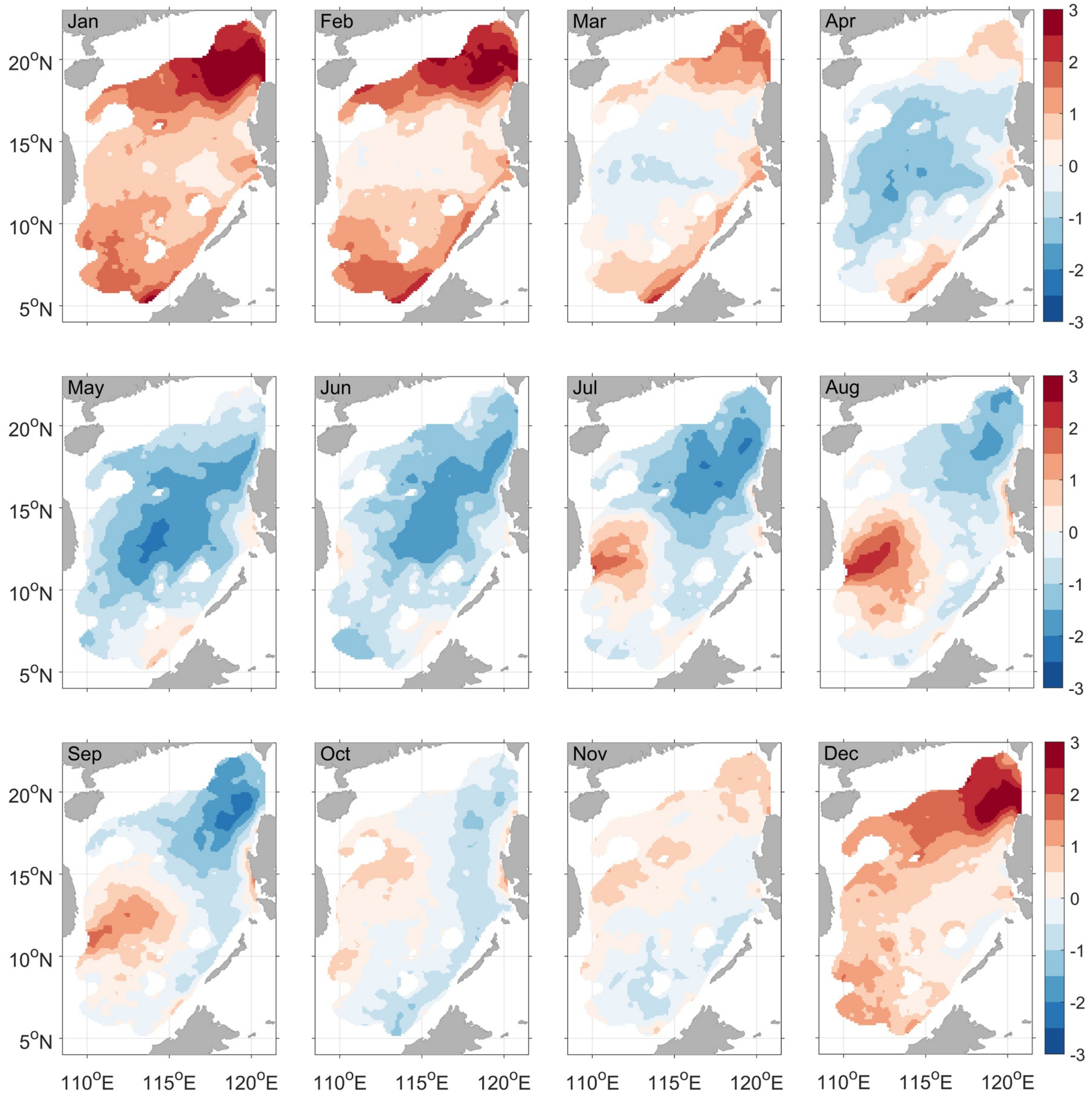

Figure 10. The spatial distribution and temporal evolution of the first principal dynamics. We flattened spatiotemporal characteristics $E \in R^{T \times N \times K}$ into $E^{\text{flatten}} \in R^{TN \times K}$ and performed PCA. We obtained $E^{\text{PCA}} \in R^{T \times N \times K}$ by reshaping principal components of E^{flatten} into the original shape. $E_1^{\text{PCA}} \in R^{T \times N}$ is the first component of E^{PCA} in the third dimension. We display E_1^{PCA} normalized by its standard deviation.

Chl_a observations. The observations' main variabilities, that is, the principal dynamics and dominant EOF modes, naturally correspond to one another.

We found that the first EOF mode of Chl_a was primarily active near the Luzon Strait around January and the second mode dominated near Vietnam around August. The spatiotemporal characteristics, $E_1^{\text{PCA}} \in R^{T \times N}$, of the first principal dynamics coincided with the spatial patterns of the first two EOF modes (denoted C_1, C_2) for these regions and months (bottom of Figures 11a and 11b, and left two columns in Figure 11c). We calculated the temporal trends of the first two EOF modes by averaging their time series within each month. The trends aligned with the temporal trends of the first principal dynamics, which is E_1^{PCA} averaged within Region 1 and Region 4 of the partition, respectively (top rows in Figures 11a and 11b). Furthermore, for month t, we performed linear regression of t-th component of E_1^{PCA} in the first dimension on spatial patterns of the first 2, 5, 10, 20, and 50 EOF modes of Chl_a. Results (right column in Figure 11c) showed an increased correlation as the number of EOF

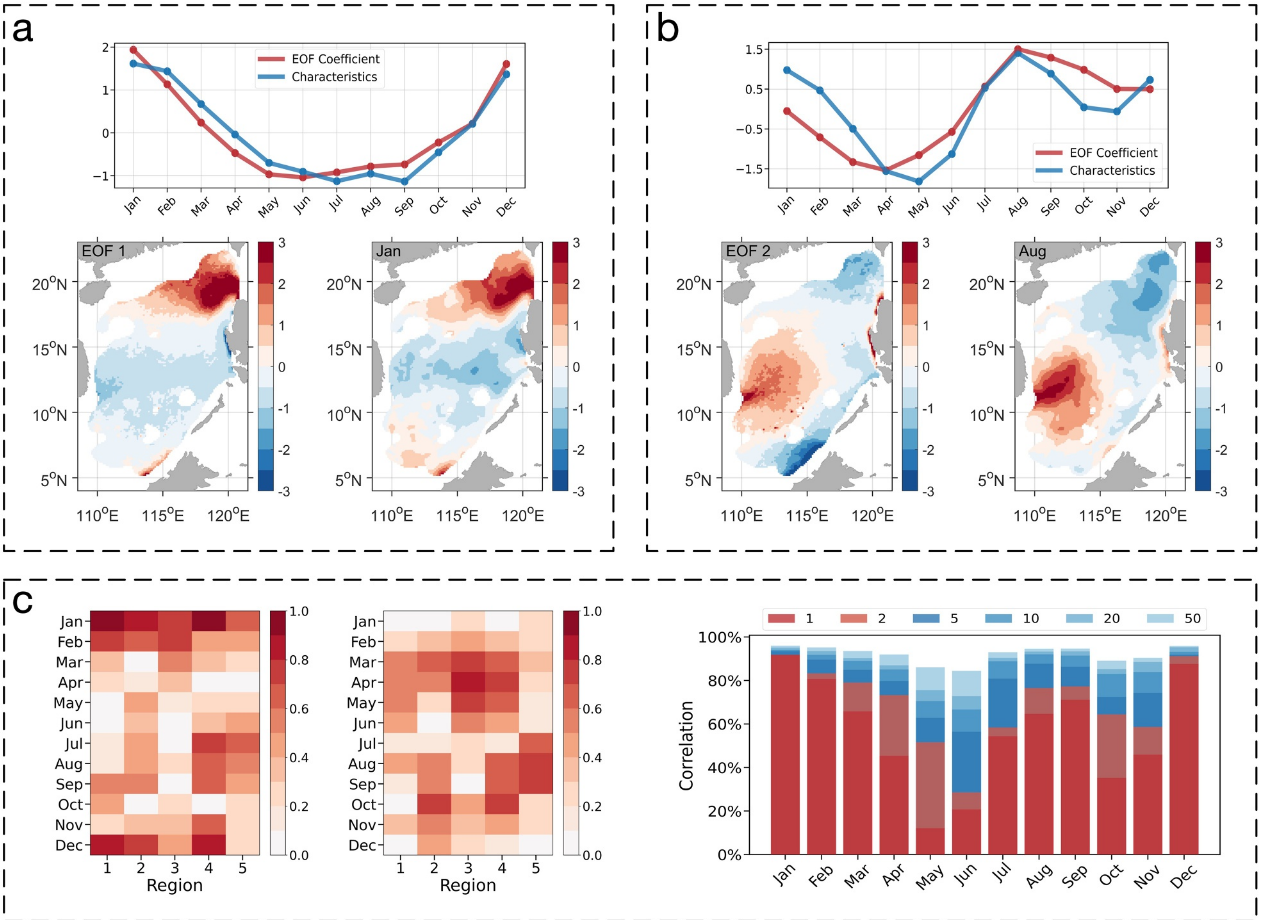

Figure 11. (a) C_1 (left in the bottom) and E_1^{PCA} averaged within Region 1 in January (right in the bottom) and their temporal trends (top). (b) C_2 (left in the bottom) and E_1^{PCA} averaged within Region 4 in August (right in the bottom) and their temporal trends (top). (c) The absolute correlation between C_1 and E_1^{PCA} (left) and between C_2 and E_1^{PCA} (middle) across regions and months. The correlation between E_1^{PCA} and the fitted E_1^{PCA} for regressions of E_1^{PCA} on spatial patterns of the first 1, 2, 5, 10, 20, and 50 EOF modes of Chl_a across months (right).

modes increases. The intrinsic connection between E_1^{PCA} and dominant EOF modes of Chl_a demonstrated the interpretable mechanism by which STD-Hunter effectively accommodates heterogeneous dynamics.

The hyperparameter K, the number of subspaces to accommodate the heterogeneous dynamics of Chl_a across space and time, is a key parameter in the STD-Hunter model. We provide an approach to identify the appropriate K. We trained STD-Hunter with a large $K = 10$, and obtained spatiotemporal characteristics E. We performed the PCA transformation to E and visualized the variance explained by each PC component (Figure 12). The first 4 PC components explained the 98.5% of variance in E, indicating that $K = 4$ was sufficient to recover the majority of the parameter tensor P.

6. Discussion and Conclusions

Using the marginal SCS in the western Pacific Ocean as the experimental region, we showed that STD-Hunter is a novel machine learning approach that facilitates long-term and large-scale prediction of highly variable Chl_a. Extensive experiments have demonstrated the superior performance of STD-Hunter in predicting spatiotemporal trends of highly variable Chl_a across various prediction horizons. Notably, STD-Hunter substantially improves predictions for regions with complex dynamics.

STD-Hunter manages heterogeneous data with a flexible module. This core module generates parameters using spatiotemporal characteristics. We substantiated the core module's effectiveness and necessity in its mechanism analysis and in the ablation study. In addition, the core module optimizes the balance between effectively utilizing

10.1029/2026JH001314

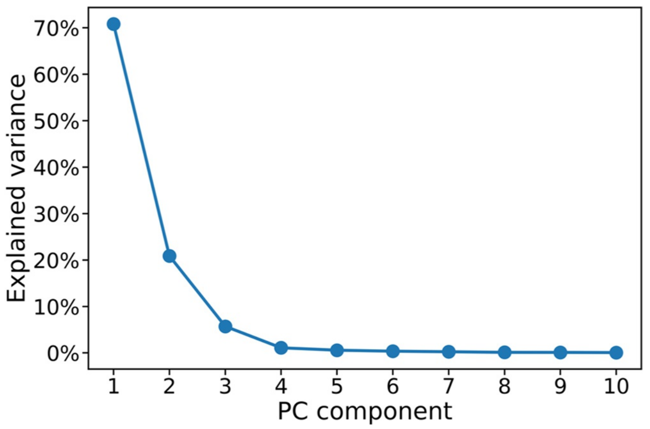

Figure 12. The explained variance of each principal component of E for $K = 10$. Specifically, we flattened spatiotemporal characteristics $E \in R^{T \times N \times K}$ of a trained STD-Hunter model with $K = 10$ into $E^{\text{flatten}} \in R^{TN \times K}$ and performed PCA.

data and preserving idiosyncratic dynamics by leveraging a low-rank approximation of the parameter tensor, which substantially mitigates the limitations of conventional approaches. The characteristics can either be learned directly from the data or specified a priori due to the nature of the tasks. In addition, the parameters to be generated are flexibly aligned with the underlying model architecture. This module is readily adaptable to incorporate into a wide range of tasks and machine learning models for enhancement.

STD-Hunter offers an interpretable data-driven approach for obtaining characteristics associated with the processes in the ocean. By substantializing veiled heterogeneity into spatiotemporal characteristics, E, we demonstrated that STD-Hunter builds a bridge between the underlying dynamics and the heterogeneous Chl_a observations we used in this study. The partition based on E corresponds well to the complex dynamics for the distribution of Chl_a in the SCS. E^{PCA} characterizes the spatial distribution and temporal evolution of each principal dynamics, effectively disentangling the multifaceted heterogeneity. The intrinsic connections between principal dynamics and Chl_a's EOF modes offer compelling validation, indicating the interpretable mechanism of STD-Hunter for predicting and diagnosing highly variable processes in the ocean.

Given that there is even more pronounced heterogeneity of the ocean worldwide, our future directions include applying STD-Hunter to predict Chl_a concentration on a global scale.

Appendix A: Benchmark Models

Our benchmark models are selected to cover different methods in space-time modeling. We employ XGBoost and LSTM as standard baselines, CNN-LSTM as a representative hybrid spatiotemporal model, and Graph WaveNet and MTGNN as two advanced spatiotemporal graph networks (STGNNs). We provided the description and the hyperparameter settings (tuned on the validation set) of the benchmark models.

1. XGBoost (Extreme Gradient Boosting): a gradient boosting algorithm that combines the predictions of multiple sequentially constructed decision trees, with each subsequent tree attempting to correct the errors made by the previous ones (Chen & Guestrin, 2016). The hyperparameters are set as eta = 0.1, n_estimators = 100, max_depth = 6.
2. LSTM (Long Short-Term Memory): a type of recurrent neural network (RNN) architecture with a gating mechanism to selectively remember and forget information (Hochreiter, & Schmidhuber, 1997). The number of features in the hidden state is 16 and the number of recurrent layers is 1.
3. CNN-LSTM: utilizes 1D CNN to extract temporal features and LSTM to combine the extracted features (Barzegar et al., 2020). The network architecture is: Convolutional layer (32 filters + 2 filter size + ELU activation + padding 'same') + Dropout layer (0.001) + LSTM layer (32 neurons + ELU activation) + Dropout layer (0.001) + Flatten + Dense layer (1 neuron).
4. Graph WaveNet: an STGNN model that learns an adaptive dependency matrix, and adds gate mechanism, skip connection, and residual connection into the original TCN (Wu et al., 2019). The hyperparameters are set as: the adjacency matrix is $\tilde{A}^{\text{space}}$, nhid = 16, learning_rate = 0.01, dropout = 0.3, weight_decay = 0.00001, epochs = 100.
5. MTGNN: a general STGNN framework featuring a graph learning module, mix-hop propagation layer, and dilated inception layer to capture the spatial and temporal dependencies (Wu et al., 2020). The hyperparameters are set as: the adjacency matrix is $\tilde{A}^{\text{space}}$, dropout = 0.3, dilation_exponential = 1, conv_channels = 16, residual_channels = 16, skip_channels = 16, end_channels = 16, layers = 3, learning_rate = 0.001, weight_decay = 0.0001, epochs = 100.

The hyperparameter settings for the STD-Hunter are as follows: dropout = 0.5, channels = 16 (number of feature channels after 1×1 Conv), layer = 3 (number of stacked temporal convolutional layers), learning_rate = 0.001, weight_decay = 1e−6, $K = 4$ (dimension of the task's characteristics), and reg = 1e−5 (coefficient for the two smoothness regularization terms).

Appendix B: Error Decomposition

We provided an illustration of error decomposition. In the training period, locations 1 and 2 are not optimized simultaneously. In the test period, they achieve the optimal test error ($error_{1,optimal}, error_{2,optimal}$) in different epochs ($epoch_{1,optimal}, epoch_{2,optimal}$). The model achieves the optimal error ($error_{suboptimal}$) at $epoch_{suboptimal}$. However, two locations achieve a suboptimal error ($error_{1,suboptimal}$ and $error_{2,suboptimal}$) at this epoch. The potential optimal error of the model is:

$$\text{error}_{\text{optimal}} = \frac{1}{2}(\text{error}_{1,\text{optimal}} + \text{error}_{2,\text{optimal}}).$$

The gap between $error_{optimal}$ and $error_{suboptimal}$ is:

$$\begin{aligned}\text{error}_{\text{suboptimal}} - \text{error}_{\text{optimal}} &= \frac{1}{2}\Big[(\text{error}_{1,\text{suboptimal}} - \text{error}_{1,\text{optimal}}) \\ &\quad + (\text{error}_{2,\text{suboptimal}} - \text{error}_{2,\text{optimal}})\Big] \\ &= \frac{1}{2}(\text{dispersion}_1 + \text{dispersion}_2) \\ &= \text{dispersion}.\end{aligned}$$

In the validation period, epoch is the selected epoch that minimizes the validation loss. At epoch, the test period error is error, which is the real model error. The gap between error and $error_{suboptimal}$ is the mismatch term:

$$\text{mismatch} = \text{error} - \text{error}_{\text{suboptimal}}.$$

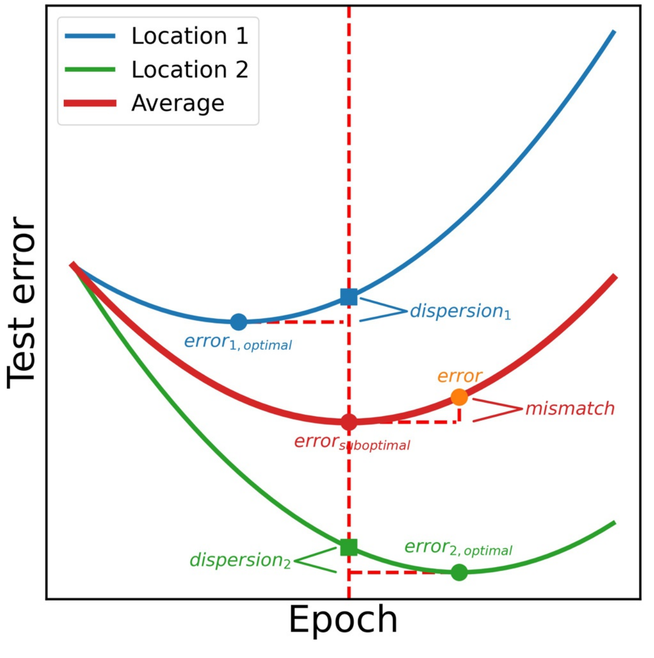

Figure B1. Illustration of error decomposition. $\text{error} = \text{error}_{\text{suboptimal}} + \text{mismatch} = \frac{1}{2}(\text{error}_{1,\text{suboptimal}} + \text{error}_{2,\text{suboptimal}}) + \text{mismatch} = \frac{1}{2}(\text{error}_{1,\text{optimal}} + \text{error}_{2,\text{optimal}}) + \frac{1}{2}(\text{dispersion}_1 + \text{dispersion}_2) + \text{mismatch} = \text{error}_{\text{optimal}} + \text{dispersion} + \text{mismatch}$.

Combing the above equations, we have:

$$\text{error} = \text{error}_{\text{optimal}} + \text{dispersion} + \text{mismatch}.$$

Figure B1.

Appendix C: Sensitivity to Winsorization

We perform winsorization with different thresholds (0%, 5%, 10%, and 15%) and report the results of 10 different runs in Figure C1. Figure C1a shows that moderate winsorization generally stabilizes model performance by mitigating outlier noise. Figure C1b displays the improvements of STD-Hunter relative to MTGNN, demonstrating that the relative performance gain of our model over the top baselines remains significant across all scenarios.

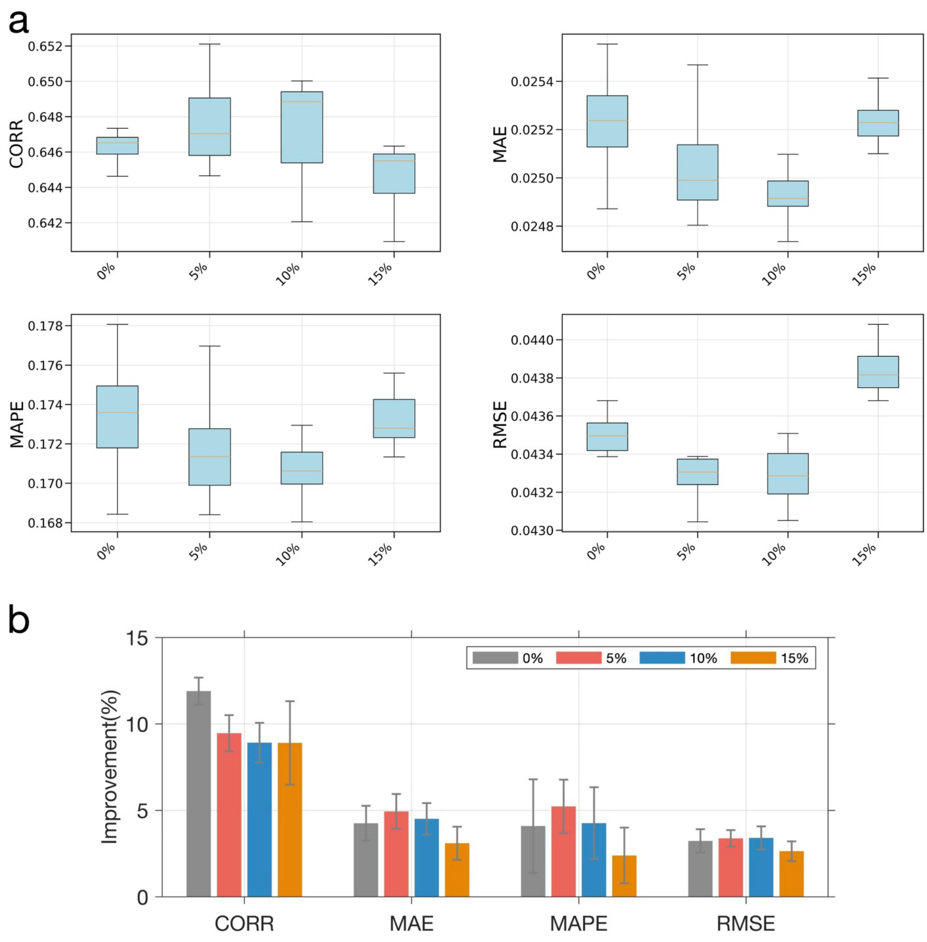

Figure C1. Sensitivity analysis across various winsorization thresholds over 10 independent runs. (a) Predictive performance of STD-Hunter. (b) Relative performance gains of STD-Hunter over the MTGNN baseline.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The chlorophyll-a (Chl_a) concentration and sea surface temperature (SST) data are from the MODIS Aqua projects (Esaias et al., 1998). The historical observations of Chl_a are from the OC-CCI data set (Sathyendranath et al., 2023). The sea level anomalies (SLA) and the absolute geostrophic velocity ($\vec{U}_{\text{ocean}}$) are from AVISO DT global map products (Copernicus Climate Change Service & Climate Data Store, 2018). The atmospheric related drivers (wind speed, zonal and meridional wind) are from the ERA5 products (Hersbach et al., 2023). The data set used for prediction is available in Zhang (2026), and the source code for STD-Hunter and the benchmark models can be found in Zhang et al. (2026).

Acknowledgments

Our work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/P-601/23-N). The General Research Fund (GRF 16310724) of the Hong Kong Research Grants Council also supported this work. The data for this study can be obtained following Data Availability Statement below.

References

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

Barzegar, R., Aalami, M. T., & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, *34*(2), 415–433. https://doi.org/10.1007/s00477-020-01776-2

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography*, *42*(1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, *619*(7970), 533–538. https://doi.org/10.1038/s41586-023-06185-3

Boyce, D. G., Lewis, M. R., & Worm, B. (2010). Global phytoplankton decline over the past century. *Nature*, *466*(7306), 591–596. https://doi.org/10.1038/nature09268

Brando, V., Dekker, A., Marks, A., Qin, Y., & Oubelkheir, K. (2006). Chlorophyll and suspended sediment assessment in a macrotidal tropical estuary adjacent to the Great Barrier Reef: Spatial and temporal assessment using remote sensing. *Cooperative Research Centre for Coastal Zone, Estuary & Waterway Management Technical Report*, *74*.

Chen, L., Zhong, X., Li, H., Wu, J., Lu, B., Chen, D., et al. (2024). A machine learning model that outperforms conventional global subseasonal forecast models. *Nature Communications*, *15*(1), 6425. https://doi.org/10.1038/s41467-024-50714-1

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Cho, H., Choi, U. J., & Park, H. (2018). Deep learning application to time-series prediction of daily chlorophyll-a concentration. *WIT Transactions on Ecology and the Environment*, *215*, 157–163.

Copernicus Climate Change Service, Climate Data Store. (2018). Sea level gridded data from satellite observations for the global ocean from 1993 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.4c328c78

Corradini, F., Gerosa, F., Gori, M., Lucheroni, C., Piangerelli, M., & Zannotti, M. (2024). A systematic literature review of spatio-temporal graph neural network models for time series forecasting and classification. *arXiv preprint arXiv:2410.22377*.

Du, C., Gan, J., Hui, C. R., Lu, Z., Zhao, X., Roberts, E., & Dai, M. (2020). Dynamics of dissolved inorganic carbon in the South China Sea: A modeling study. *Progress in Oceanography*, *186*, 102367. https://doi.org/10.1016/j.pocean.2020.102367

Esaias, W. E., Abbott, M. R., Barton, I., Brown, O. B., Campbell, J. W., Carder, K. L., et al. (1998). An overview of MODIS capabilities for ocean science observations. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1250–1265. https://doi.org/10.1109/36.701076

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/essd-14-4811-2022

Gan, J., Kung, H., Cai, Z., Liu, Z., Hui, C., & Li, J. (2022). Hotspots of the stokes rotating circulation in a large marginal sea. *Nature Communications*, *13*(1), 2223. https://doi.org/10.1038/s41467-022-29610-z

Gan, J., Liu, Z., & Hui, C. R. (2016). A three-layer alternating spinning circulation in the South China Sea. *Journal of Physical Oceanography*, *46*(8), 2309–2315. https://doi.org/10.1175/jpo-d-16-0044.1

Gan, J., Lu, Z., Cheung, A., Dai, M., Liang, L., Harrison, P. J., & Zhao, X. (2014). Assessing ecosystem response to phosphorus and nitrogen limitation in the P earl R iver plume using the R egional O cean M odeling S ystem (ROMS). *Journal of Geophysical Research: Oceans*, *119*(12), 8858–8877. https://doi.org/10.1002/2014jc009951

Gan, J., Lu, Z., Dai, M., Cheung, A. Y., Liu, H., & Harrison, P. (2010). Biological response to intensified upwelling and to a river plume in the northeastern South China Sea: A modeling study. *Journal of Geophysical Research*, *115*(C9). https://doi.org/10.1029/2009jc005569

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C., Bopp, L., et al. (2020). Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Frontiers in Marine Science*, *7*, 571720. https://doi.org/10.3389/fmars.2020.571720

Hauser, D., Tourain, C., Hermozo, L., Alraddawi, D., Aouf, L., Chapron, B., et al. (2020). New observations from the SWIM radar on-board CFOSAT: Instrument validation and ocean wave measurement assessment. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(1), 5–26. https://doi.org/10.1109/tgrs.2020.2994372

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., & Thépaut, J.-N. (2023). ERA5 hourly data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly journal of the royal meteorological society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hofmann, E. E., Cahill, B., Fennel, K., Friedrichs, M. A., Hyde, K., Lee, C., et al. (2011). Modeling the dynamics of continental shelf carbon. *Annual Review of Marine Science*, *3*(1), 93–122. https://doi.org/10.1146/annurev-marine-120709-142740

Lacroix, F., Ilyina, T., Mathis, M., Laruelle, G. G., & Regnier, P. (2021). Historical increases in land-derived nutrient inputs may alleviate effects of a changing physical climate on the oceanic carbon cycle. *Global Change Biology*, *27*(21), 5491–5513. https://doi.org/10.1111/gcb.15822

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. (2017). Temporal convolutional networks for action segmentation and detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 156–165).

10.1029/2026JH001314

Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal convolutional networks: A unified approach to action segmentation. In *European conference on computer vision* (pp. 47–54). Springer International Publishing.

Li, X., Sha, J., & Wang, Z. L. (2018). Application of feature selection and regression models for chlorophyll-a prediction in a shallow lake. *Environmental Science and Pollution Research*, *25*(20), 19488–19498. https://doi.org/10.1007/s11356-018-2147-3

Liu, N., Chen, S., Cheng, Z., Wang, X., Xiao, Y., Xiao, L., & Liu, S. (2022). Long-term prediction of sea surface chlorophyll-a concentration based on the combination of spatio-temporal features. *Water Research*, *211*, 118040.

Long, M. C., Moore, J. K., Lindsay, K., Levy, M., Doney, S. C., Luo, J. Y., et al. (2021). Simulations with the marine biogeochemistry library (MARBL). *Journal of Advances in Modeling Earth Systems*, *13*(12), e2021MS002647. https://doi.org/10.1029/2021ms002647

Lu, Z., & Gan, J. (2023). A modeling study of nutrient transport and dynamics over the northern slope of the South China Sea. *Journal of Geophysical Research: Oceans*, *128*(5), e2022JC019225. https://doi.org/10.1029/2022jc019225

Lu, Z., Gan, J., Dai, M., Zhao, X., & Hui, C. R. (2020). Nutrient transport and dynamics in the South China Sea: A modeling study. *Progress in Oceanography*, *183*, 102308. https://doi.org/10.1016/j.pocean.2020.102308

Lu, Z., Yu, L., & Gan, J. (2022). External and internal forcings for hypoxia formation in an Urban harbour in Hong Kong. *Frontiers in Marine Science*, *9*, 858715. https://doi.org/10.3389/fmars.2022.858715

Matear, R. J. (1995). Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., et al. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. https://doi.org/10.1038/s41586-024-07145-1

Ning, X. R., Chai, F., Xue, H. J., Cai, Y., Liu, C., & Shi, J. (2004). Physical-biological oceanographic coupling influencing phytoplankton and primary production in the South China Sea. *Journal of Geophysical Research*, *109*(C10). https://doi.org/10.1029/2004jc002365

Palacz, A. P., Xue, H., Armbrecht, C., Zhang, C., & Chai, F. (2011). Seasonal and inter-annual changes in the surface chlorophyll of the South China Sea. *Journal of Geophysical Research*, *116*(C9), C09015. https://doi.org/10.1029/2011jc007064

Park, Y., Cho, K. H., Park, J., Cha, S. M., & Kim, J. H. (2015). Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea. *Science of the Total Environment*, *502*, 31–41. https://doi.org/10.1016/j.scitotenv.2014.09.005

Sathyendranath, S., Brewin, R. J., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., et al. (2019). An ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change initiative (OC-CCI). *Sensors*, *19*(19), 4285. https://doi.org/10.3390/s19194285

Sathyendranath, S., Jackson, T., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., & Platt, T. (2023). ESA ocean colour climate change initiative (Ocean_Colour_cci): Monthly climatology of global ocean colour data products at 4km resolution, version 6.0 [Dataset]. *NERC EDS Centre for Environmental Data Analysis*, *2025–07–21*. https://catalogue.ceda.ac.uk/uuid/690fdf8f229c4d04a2aa68de67beb733

Shao, Z., Wang, F., Xu, Y., Wei, W., Yu, C., Zhang, Z., & Cheng, X. (2024). Exploring progress in multivariate time series forecasting: Comprehensive benchmarking and heterogeneity analysis. *IEEE Transactions on Knowledge and Data Engineering*.

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: A survey. *Big Data*, *9*(1), 3–21. https://doi.org/10.1089/big.2020.0159

Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, *12*, 1.

Wang, Y., Wu, H., Dong, J., Liu, Y., Wang, C., Long, M., & Wang, J. (2024). Deep time series models: A comprehensive survey and benchmark. *arXiv preprint arXiv:2407.13278*.

Wright, R. M., Le Quéré, C., Buitenhuis, E., Pitois, S., & Gibbons, M. J. (2021). Role of jellyfish in the plankton ecosystem revealed using a global ocean biogeochemical model. *Biogeosciences*, *18*(4), 1291–1320. https://doi.org/10.5194/bg-18-1291-2021

Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 753–763).

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*.

Xu, W., Wang, G., Jiang, L., Cheng, X., Zhou, W., & Cao, W. (2020). Spatiotemporal variability of surface phytoplankton carbon and carbon-to-chlorophyll a ratio in the South China Sea based on satellite data. *Remote Sensing*, *13*(1), 30. https://doi.org/10.3390/rs13010030

Yajima, H., & Derot, J. (2018). Application of the random forest model for chlorophyll-a forecasts in fresh and brackish water bodies in Japan, using multivariate long-term databases. *Journal of Hydroinformatics*, *20*(1), 206–220. https://doi.org/10.2166/hydro.2017.010

Yan, J., Mu, L., Wang, L., Ranjan, R., & Zomaya, A. Y. (2020). Temporal convolutional networks for the advance prediction of ENSO. *Scientific Reports*, *10*(1), 8055. https://doi.org/10.1038/s41598-020-65070-5

Yang, Y., Jin, M., Wen, H., Zhang, C., Liang, Y., Ma, L., & Wen, Q. (2024). A survey on diffusion models for time series and spatio-temporal data. *arXiv preprint arXiv:2404.18886*.

Zhang, F. (2026). Taming the heterogeneous dynamics of Oceanic Chlorophyll-a concentration with a novel deep learning model to improve prediction (v1.0) [Code]. *Zenodo*. https://doi.org/10.5281/zenodo.21204043

Zhang, F., Kung, H., Zhang, F., Wang, Z., Yang, C., & Gan, J. (2026). Taming the heterogeneous dynamics of Oceanic Chlorophyll-a concentration with a novel deep learning model to improve prediction [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.21203532

Zhang, Y., Long, M., Chen, K., Xing, L., Jin, R., Jordan, M. I., & Wang, J. (2023). Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, *619*(7970), 526–532. https://doi.org/10.1038/s41586-023-06184-4

Zheng, L., Wang, H., Liu, C., Zhang, S., Ding, A., Xie, E., et al. (2021). Prediction of harmful algal blooms in large water bodies using the combined EFDC and LSTM models. *Journal of Environmental Management*, *295*, 113060. https://doi.org/10.1016/j.jenvman.2021.113060