

A Spatio-Temporal Attention Graph Convolutional Networks for Sea Surface Temperature Prediction

Desheng Chen, Jiabao Wen*, Caiyun Lv

School of Electrical and Information Engineering, Tianjin University, Tianjin (P.R. China)

Received 16 June 2022 | Accepted 10 February 2023 | Early Access 23 February 2023



ABSTRACT

Sea surface temperature (SST) is an important index to detect ocean changes, predict SST anomalies, and prevent natural disasters caused by abnormal changes, dynamic variation of which have a profound impact on the whole marine ecosystem and the dynamic changes of climate. In order to better capture the dynamic changes of ocean temperature, it's vitally essential to predict the SST in the future. A new spatio-temporal attention graph convolutional network (STAGCN) for SST prediction was proposed in this paper which can capture spatial dependence and temporal correlation in the way of integrating gated recurrent unit (GRU) model with graph convolutional network (GCN) and introduced attention mechanism. The STAGCN model adopts the GCN model to learn the topological structure between ocean location points for extracting the spatial characteristics from the ocean position nodes network. Besides, capturing temporal correlation by learning dynamic variation of SST time series data, a GRU model is introduced into the STAGCN model to deal with the prediction problem about long time series, the input of which is the SST data with spatial characteristics. To capture the significance of SST information at different times and increase the accuracy of SST forecast, the attention mechanism was used to obtain the spatial and temporal characteristics globally. In this study, the proposed STAGCN model was trained and tested on the East China Sea. Experiments with different prediction lengths show that the model can capture the spatio-temporal correlation of regional-scale sea surface temperature series and almost uniformly outperforms other classical models under different sea areas and different prediction levels, in which the root mean square error is reduced by about 0.2 compared with the LSTM model.

KEYWORDS

Data Forecast, Gated Recurrent Unit, Graph Convolutional Network, Sea Surface Temperature, Spatiotemporal Attention Graph Convolutional Network.

DOI: 10.9781/ijimai.2023.02.011

I. INTRODUCTION

THE dynamics of the oceans, which make up about two-thirds of the planet, have an extremely important impact on climate, marine ecology, and the lives of the people around them. Sea surface temperature (SST) is an important index to detect ocean changes, predict SST anomalies, and prevent natural disasters caused by abnormal changes. Therefore, it's significant to predict the dynamic change of SST in the future. Moreover, SST has played an indispensable role in the ocean-atmosphere interaction, that is, the exchange of matter, energy, and momentum between the ocean and the atmosphere [1] [2] [3]. As a result, changes in SST have incalculable impacts on global climate and marine ecosystem [4] [5] [6] [7]. Besides, SST predictions also has implications for applications related to the ocean, such as weather forecasting, fisheries, and marine environmental protection. Therefore, it's critically necessary to predict dynamic changes of SST in the future to help people identify and prevent severe weather events such as drought in advance [8] [9], and it's also of great significance for scientific research and applications [10]. However, due to the influence

of many complex factors, such as sea surface heat flow, radiation, and solar wind, the prediction of SST is quite indefinite and challenging.

In recent years, SST prediction methods have been widely applied and attracted much attention further become an attractive field of marine research. Three kinds of methods are generally used to predict SST including the numerical method based on the mathematical model, the data-driven method using the historical model to predict SST in the future, and the method combining the two methods [11]. The numerical methods generally use kinetic and thermodynamic equations to exposit the dynamic changes of SST and then solve a series of differential equations which are difficult to solve due to they are usually sophisticated and require a large amount of computation. The data-driven method is mainly used to predict the future SST value from the perspective of data. This method builds the model by learning the relationships and patterns from the historical SST observation data and further uses the learned relationships model to approximate the future SST data. The data-driven method is less complicated than the numerical method and is suitable for the prediction of SST in high-resolution areas. The data-driven method mainly predicts the future SST from the perspective of statistical data analysis, machine learning, and artificial intelligence algorithms. Among them, statistical data analysis techniques primarily contain the Markov model [12] [13], Empirical canonical correlation analysis [14] and regression model

* Corresponding author.

E-mail address: Wen_Jiabao@tju.edu.cn

[15] [16], etc. Classical machine learning methods including linear regression, support vector machine (SVM) [17], nonlinear regression model [18], and artificial neural network [19] are used to forecast future SST. Support vector machine (SVM) is a generalized linear classifier that classifies data based on supervised learning. Particle swarm optimization (PSO) algorithm is a random and parallel optimization algorithm based on population. Those two kinds of artificial intelligence methods are commonly used in SST prediction [17] [20]. The numerical method and machine learning method can also be combined [11] to better predict SST, but the prediction effect is similar to that using the numerical method.

With the continuous development and innovation of deep learning, the deep learning method has been mostly used in SST prediction due to its powerful ability to learn and model the relationship between data [21] [22] [23] [24]. Recurrent neural network (RNN) can effectively deal with time series prediction problems, but serious gradient vanishing or outbreak problems will occur when processing long time series data. As a variant of the recurrent neural network, a long short-term memory (LSTM) network with recurrent structure and gating mechanism is proposed to solve the long-term time dependence problem, which can remember longer time series information and obtain better prediction results [25]. In order to simplify the complex structure of LSTM network, a gate recurrent unit (GRU) [26] with relatively unsophisticated gate structure was proposed. As an improved variant of the LSTM network, the GRU model not only retains the advantage of LSTM in long-term series memory but also has high computational efficiency, which alleviated the phenomenon of network overfitting and underfitting. Those model have been widely used in ocean surface temperature prediction [27] proposed a full-connected LSTM (FC-LSTM), which is composed of LSTM layer and Full Connected layer. [26] designs an adaptive mechanism based on deep learning and attention mechanism to predict SST, which uses GRU encoder-decoder to obtain the static change of SST and apply dynamic influence link to acquire the dynamic variation for realizing the long-term prediction of future SST. [9] proposes an integrated learning model (LSTM-Adaboost) that combines the deep LSTM neural network with the Adaptive Boosting (AdaBoost) algorithm to predict the daily SST in the short and medium-term. Feng et al used time-domain convolutional network to achieve short-term small-scale SST prediction of the Indian Ocean [28]. Han et al. used convolutional neural network method to achieve regional prediction of Sea surface temperature, sea surface height and ocean salinity in the Pacific [29].

However, although these SST prediction models have achieved a good prediction effect, they only consider the time correlation but ignore the spatial dependency, so they cannot achieve high accuracy in predicting the dynamic changes of SST sequence data. Beside, the association structures constructed when capturing the spatial influence of adjacent nodes on the central node are not all standardized grid structures. For example, topology structures based on spatial association can be constructed when missing values exist. Therefore, in order to capturing spatial dependencies from complex topologies, the GCN was applied to obtain the spatial dependence from the SST series data of the ocean location points, an original SST prediction method named spatio-temporal attention graph convolutional network (STAGCN) based on ocean location points network was proposed in this paper. Specifically, the GCN is applied to capture spatial correlation from the ocean positions network with the topological structure. The GRU is used to capturing the temporal dependence from the dynamic changes of the SST time series data. In addition, the STAGCN model introduces an attention mechanism to learn global correlation, adjust and integrate global temporal information of SST for realizing accurate SST prediction tasks eventually.

The contribution of this paper can be summarized in the following two aspects: (1) A STAGCN model is designed to capture the global spatial and temporal dependence simultaneously for accurate SST prediction, which combines the GCN deep learning model with the GRU learning model and introduces an attention mechanism. (2) The concept of a graph is applied to the field of SST prediction, and the topology structure network of ocean location points graph is constructed to obtain the spatial characteristics from SST time series by GCN model. Using 38-year time-series satellite data from some areas of the East China Sea, the experiments show that the STAGCN model can achieve preferable prediction results than the GRU model and the GCN model demonstrating that the STAGCN model has the ability to capture both time and space correlation from SST series data simultaneously, and can achieve desirable prediction effect for the short-term prediction of future SST.

The rest of the paper is organized as follows. Section II proposes the novel STAGCN model for SST prediction in detail. Section III describes the experiments and analyzes the results. Finally, Section IV gives the conclusion of the paper.

II. METHOD

A. Problem Clarification

In this study, the prediction of SST is to predict the sea surface temperature within a certain time in the future according to the historical SST time-series information.

The undirected graph $G = (V, E)$ with no weights was used to describe a topological network composed of oceanic observation points, and each location represents a node in the graph, where $V = v_1, v_2, \dots, v_N$ means all oceanic positions that correspond to the N vertices of the graph. E represents the correlation between ocean points corresponding to the edges between nodes in the graph, reflecting the connection between nodes at different positions. The connection relationship between nodes in the whole graph is represented by the adjacencies matrix A , the number of the rows and columns of which are determined by the number of nodes, and each value of which represents the connection relationship between nodes. The value of each item in the adjacency matrix A is either 0 or 1. 0 indicates that there is no direct connection between nodes and 1 expresses that there has a linkage between nodes.

The feature $X^{P \times N}$ of the node in the topology graph corresponds to the SST value of each location point on the ocean positions network. where N represents the length of the time dimension in oceanographic satellite data, and $X_t \in R^{P \times N}$ represents the SST value at time t .

The SST prediction problem can be regarded as looking for a mapping function f , which map the SST value in the historical time n to the SST value in the next T periods under the conditions of the topology diagram G of the ocean location points network and feature matrix X . Equation (1) refers to the SST prediction process:

$$(X_{t-n}, \dots, X_{t-1}, X_t) f(\bullet) [X_{t+1}, X_{t+2}, \dots, X_{t+T}] \quad (1)$$

The left side is the historical SST value with the length n , from which the model learns the variation trend of SST, and the right side of the equation is the predicted future SST value with length T under the mapping condition.

B. The STAGCN Model

To capture global dynamic changes of the SST information, a novel STAGCN model combines graph convolutional network capturing spatial correlation and gated recurrent unit obtaining temporal dependence with attention model at the same time is proposed in this paper. In the STAGCN model, a layer of GCN network can be used to

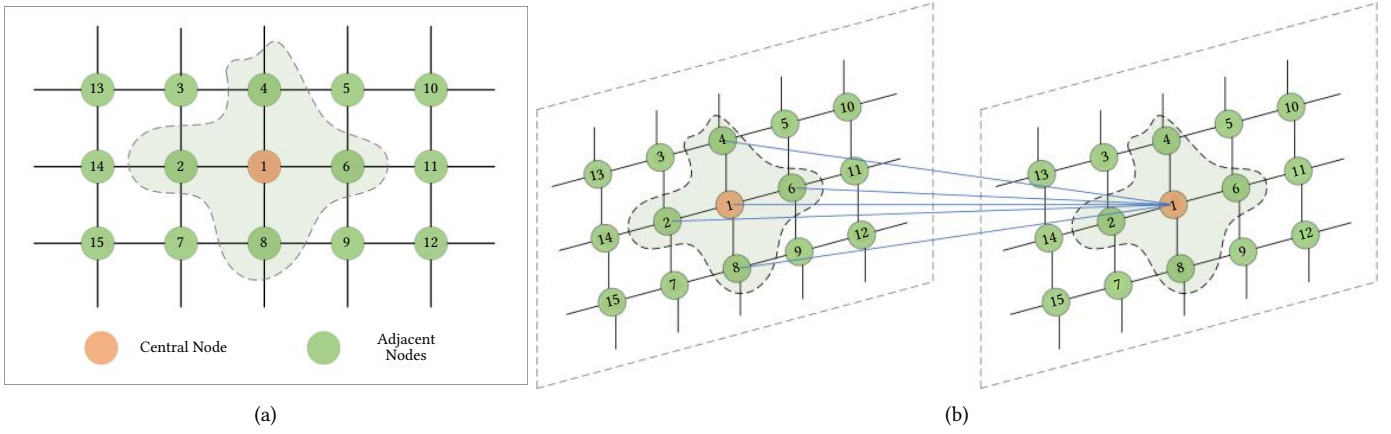


Fig. 1. Topology graph of an ocean location point (in the dotted shaded part). (a) The green nodes indicate the location points have connection relationship with the central point. (b) The topological structure between the central node 1 and its adjacent nodes is established through the GCN model to obtain the spatial features.

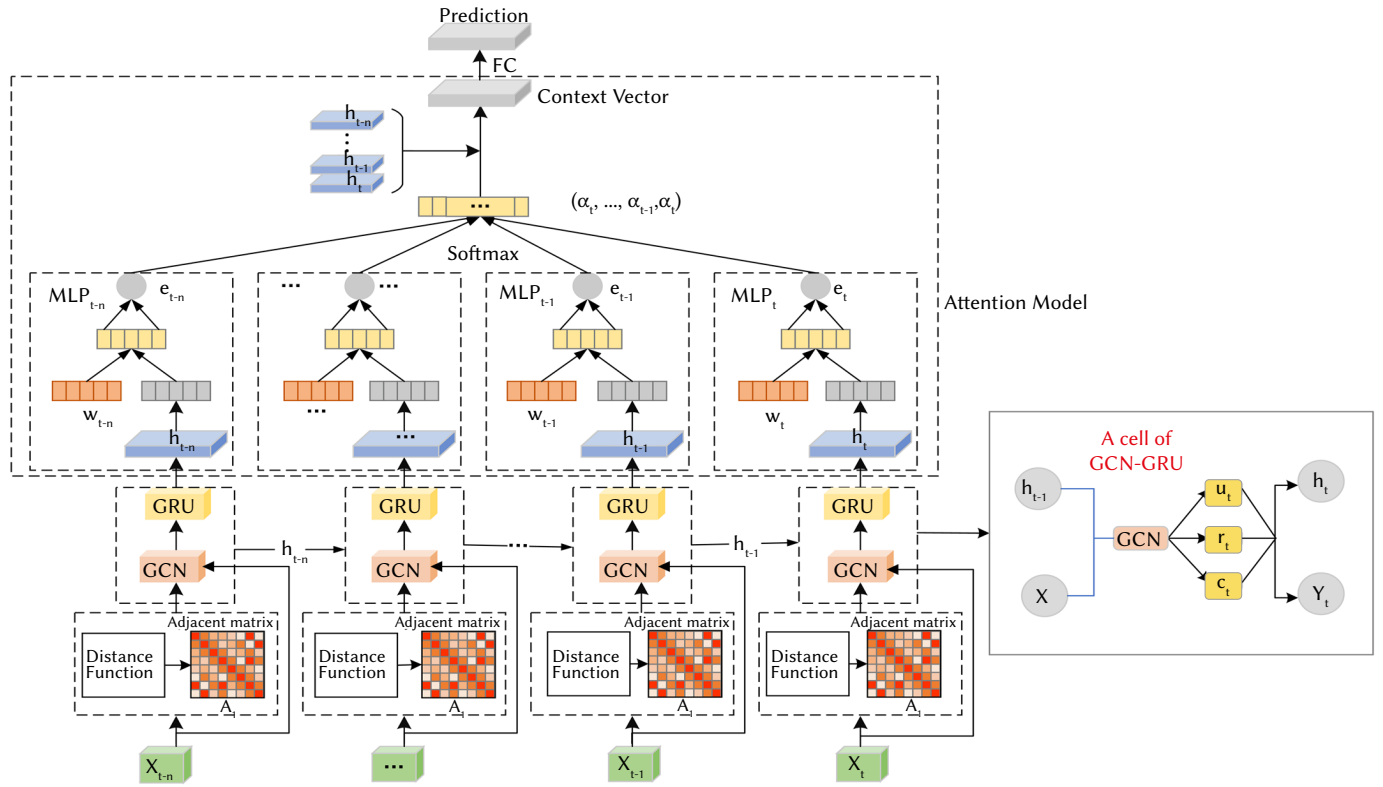


Fig. 2. The specific spatial-temporal prediction process of the STAGCN model.

obtain a better prediction effect, and Equation (2) refers to the specific convolution process:

$$f(X, A) = \sigma(\overline{A}ReLU(\overline{A}XW_0)W_1) \quad (2)$$

where $f(X, A)$ express the final output of the GCN model, X is the eigenmatrix, A means the adjacency matrix of the graph convolution. $ReLU(\cdot)$ represents an activation function used to add linear factors to improve model expression. \overline{A} is the form of the adjacent matrix after further renormalization to avoid causing gradient explosion. W_0 represents the weight that needs to be trained in the progress of the graph convolutional network, whose first dimension F means the time series length of ocean data and second dimension G means the number of neural units in the output layer.

Specifically, the area we study is ocean location points graphs composed of position points determined by longitude and latitude, and the topological connection relationship between location points can be captured by the GCN network. The shaded part of the dotted

line in Fig. 1 is a topological graph of a point in the simulated ocean location. It is assumed that the red dot 1 represents the central node in the topology diagram, and the green dots (the green dot in the shaded part of the dotted line) around it are the adjacent nodes. The interaction degree between the central node and the adjacent nodes can be obtained by GCN model. Then the GCN model captures the SST characteristic attributes of the topological structure and further acquires the spatial correlation from the location points network. The GRU model was adopted in this study to obtain the temporal dependence of SST data. The attention model is used to screen which moments of SST data are relevant, that is, to distinguish the importance of data at different moments, which improves the accuracy of the prediction and realize the SST prediction task based on the structure of the ocean location points graph. The specific spatial-temporal prediction process of the STAGCN model is shown in Fig. 2, where the distance function (Equation) was applied to calculate the adjacency matrix A_1 that represents the connection relation between the position

points, the GCN-GRU represents the process of combining the output of the GCN model and the GRU model, and the FC express the fully connected layer.

First of all, the obtained adjacency matrix and the SST feature data $X_i (i = t-n, \dots, t-1, t)$ of n historical time series were taken as input, from which the GCN model capture the spatial information. Then, The input of the GRU model is replaced by the output of the GCN model to obtain the temporal characteristics of SST data. Equation (3) to (6) refer to the update gate, reset gate, cell state and output state at time t in the STAGCN model respectively.

$$u_t = \sigma(W_u \cdot f_{gcn}(A, [h_{t-1}, X_t]) + b_u) \quad (3)$$

$$r_t = \sigma(W_r \cdot f_{gcn}(A, [h_{t-1}, X_t]) + b_r) \quad (4)$$

$$c_t = \tanh(W_c \cdot f_{gcn}(A, [(r_t * h_{t-1}), X_t]) + b_c) \quad (5)$$

$$h_t = (u_t * h_{t-1} + (1 - u_t) * c_t) \quad (6)$$

where u_t and r_t express the update gate and reset gate, h_{t-1} means the output at previous time, h_t means the state output at present time, c_t represents the information be reserved from previous moment and present time. Function $f_{gcn}(\cdot)$ express the graph convolution, W_u , W_r and W_c are the connection weights between the output of graph convolution and the previous output h_{t-1} , b_u , b_r and b_c are the corresponding thresholds. Moreover, the final hidden state information of the GRU model is used as the input of the attention model, which is used to obtain the importance of the changes information of SST series data. Finally, we get the prediction from the full connection layer. In the attention model, multi-layer perception is used as the scoring function, in which $w_i (i = t-n, \dots, t-1, t)$ is the weight matrix of multi-layer perception. The score $e_i (i = t-n, \dots, t-1, t)$ of the multi-layer perception output is brought into the Softmax function to get the attention distribution probability. The last hidden state and its weight are weighted to obtain the final context vector C .

To sum up, we propose that the STAGCN model has the ability to obtain the global spatial dependence and temporal dynamic changes which can obtain a preferable SST prediction effect. The GCN model is applied to capture spatial information by building the structure of the interrelation between the position nodes. The GRU model is used to obtain the temporal dependence from the SST series data with spatial characteristics. Moreover, the attention model captures the global variation trends of the SST information which is significant to achieve accurate SST prediction tasks.

III. EXPERIMENTS

A. Research Area and Data

Rich in natural resources, the East China Sea is the confluence of many rivers covering a wide area including China's Bohai Sea in the north and the Taiwan Strait in the south and is the strategic maritime area for China, Japan, South Korea, and other countries. Therefore, studying the dynamic changes of SST data in the East China Sea plays an extremely important role in national marine transportation and people's production and life in the surrounding countries. The research area selected in this study is the part area of the East China Sea with sea areas of $26.8755^\circ\text{N} - 32.125^\circ\text{N}$ and $123.125^\circ\text{E} - 127.125^\circ\text{E}$ covering most areas of the East China Sea and no land area, as is shown in Fig. 3. The selected area includes most of the East China Sea to facilitate the acquisition of grid-based SST data and the establishment of the topological structure of the ocean position points graph for SST prediction.

In this study, the data used in the experiments were derived from the daily Optimum interpolated SST (OISST Daily Edition 2.1) with a spatial resolution of $1/4^\circ$ in the National Oceanic and Atmospheric

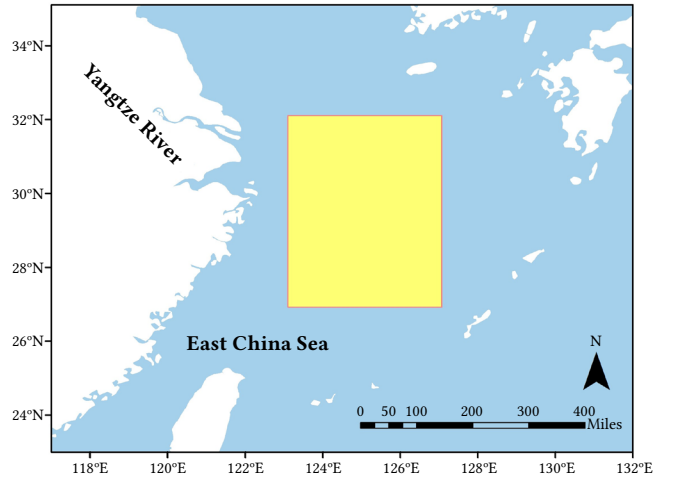


Fig. 3. Research area in the East China Sea (in yellow).

Administration (NOAA) platform. In this study, AVHRR-only daily SST data contained a total of 13,879 days SST data from 1982/01/01-2019/12/31 covered the spatial span of $26.875^\circ\text{N}-32.125^\circ\text{N}$, $123.125^\circ\text{E}-127.125^\circ\text{E}$ (the most of the east sea) are used as the experimental dataset, with a total of 22×17 position points as the study nodes.

In this study, the experimental data is composed of two parts: the adjacency matrix and the eigenmatrix respectively. The former is an adjacency matrix of size 374×374 , which depicts the spatial dependence between position points. Equation (7) refers to each value in the adjacency matrix W is derived by some scaling of the distance between each position point in the ocean anchor point network.

$$W_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right), & i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where W_{ij} is the weight of the edges in the position graph determined by the distance between position i and j (d_{ij}). σ^2 and ε are thresholds to control spacial arrangement and sparse arrangement of the adjacency matrix W , which are set to 0.1 and 0.4 respectively after testing in the experiments. The latter part of the experimental data is the eigenmatrix, which describes the dynamic changes of the SST value over time at the position nodes.

B. Experiment Setup

In order to better explain the superiority of the STAGCN model proposed in predicting SST, we chose five comparative models including the autoregressive moving average model (ARIMA), linear support vector machine model (SVR), graph convolutional network model (GCN), and gated recurrent unit model (GRU). In the experiments, we divided the dataset into the training set and test set, and the ratio of the two is 8:2. The Adam optimizer is used to train the model. The STAGCN model, GCN model, and GRU model use TensorFlow 1.5.0 (GPU version) as runtime environment during the training and testing progress. The ARIMA and SVR are respectively implemented using Statsmodels 0.12.2 and Scikit-learn 0.24.1 [31].

The real SST and the predicted SST of different nodes at time t are respectively expressed by Y_t and \hat{Y}_t . In the network training, the loss function value should be minimized as far as possible, which is beneficial for the predicted SST value of each ocean position node to be a better fit to the actual SST value. The loss function defined in the STAGCN model is shown in Equation. The first term $\| \cdot \|$ is the 2-norm of the real value and the predicted value, and the second term is the regularization term λL_{reg} with hyperparameter λ , which improved the prediction performance of the model and prevented overfitting occurrence during training.

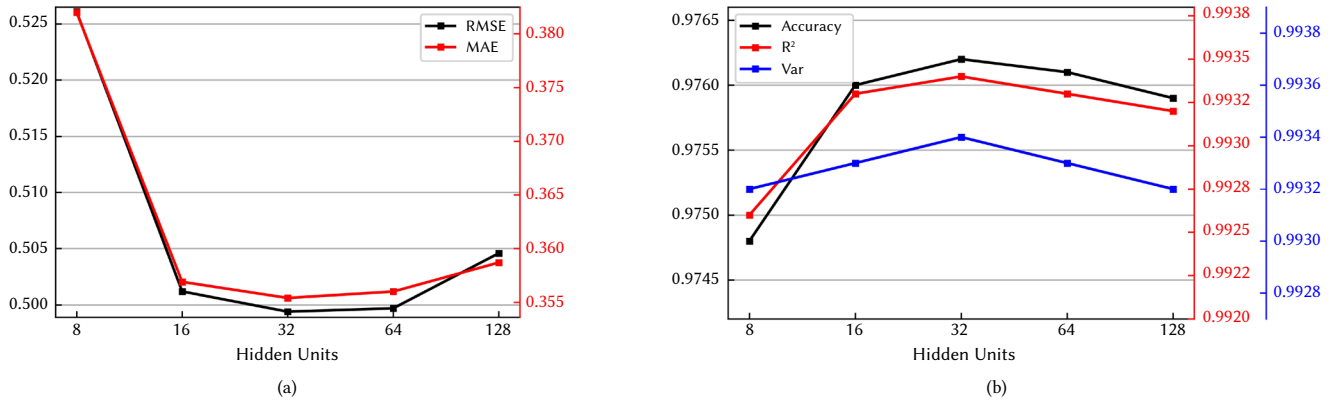


Fig. 4. Comparison of SST prediction results of the STAGCN model under different hidden units conditions. (a) Changes trend of RMSE and MAE under different hidden units conditions. (b) Variation trend in Accuracy, R^2 and Var under different hidden units conditions.

To prove the desirable prediction performance of the STAGCN model, five measurement criteria are used to compare the SST prediction performance of the proposed model with other models including root mean square error (RMSE), mean absolute error (MAE), accuracy (Accuracy), coefficient of determination (R^2) and explained variance score (Var).

$$RMSE = \sqrt{\frac{1}{TN} \sum_{t=1}^T \sum_{n=1}^N (y_n^t - \hat{y}_n^t)^2} \quad (8)$$

$$MAE = \frac{1}{TN} \sum_{t=1}^T \sum_{n=1}^N |y_n^t - \hat{y}_n^t| \quad (9)$$

$$Accuracy = 1 - \frac{\|Y - \hat{Y}\|_F}{\|Y\|_F} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{t=1}^T \sum_{n=1}^N (y_n^t - \hat{y}_n^t)^2}{\sum_{t=1}^T \sum_{n=1}^N (y_n^t - \bar{Y})^2} \quad (11)$$

$$var = 1 - \frac{\text{var}\{Y - \hat{Y}\}}{\text{var}\{Y\}} \quad (12)$$

where Equation (8) to (12) refer to the calculation process of RMSE, MAE, Accuracy, (R^2) and (Var) respectively, T means the length of the SST time series, N represents the total number of position points. TN is the number of recorded changes in temperature values at all ocean locations with a time length of T . y_n^t represents the real SST value and the \hat{y}_n^t means predicted SST value in the position point n at the moment t . The entire sets of y_n^t and \hat{y}_n^t can be defined as Y and \hat{Y} respectively. \bar{Y} represents the average value of collection Y . Five criteria are used to measure the merits of the model from the perspective of error, accuracy, and fitting degree. The F in the accuracy indicator refers to the F-norm.

In the neural network model, the setting of model parameters is crucial to the model training, such as the number of neurons in the hidden layer, whose size determines the computational complexity and predictive performance of the whole model. Therefore, in order to improve the training efficiency and the accuracy of prediction, we used different hidden units to conduct experiments on the test set and select the corresponding number of hidden units with the best predicted effect from the predicted results. We set the number of neurons in the hidden layer as the value in [8, 16, 32, 64, 128], respectively. The variation trend of error indicators under different hidden units condition as shown in 4, from which we can see that the RMSE and MAE reach the minimum value when the number of

hidden layer units is 32. At the same time, we can see from 4 that the prediction error including RMSE and MAE has a similar decreasing trend while hidden units increases from 8 to 32, and then it shows an increasing trend when hidden units exceed 32. On the contrary, the prediction accuracy has an opposite trend, rising first and then falling. The trend of the model measurement indexes shows that there is a critical value of the number of hidden layer cells in the model. When the critical value is exceeded, the complexity of the model will increase with the increase of the number of hidden layer cells, and the performance of the model will decline simultaneously. Since the prediction performance of the model has the best performance when the number of hidden units is 32, the following experiments will be carried out under the condition that the number of hidden units is set to 32.

C. Experiment Results and Discussion

We performed prediction experiments on 1 day, 7 days, 14 days, and 30 days SST values in the future and measure the performance of the proposed STAGCN model, the ARIMA model, SVR model, GCN model, and GRU model with five performance metrics. The SST prediction result of the STAGCN model is analyzed from the perspectives of prediction accuracy, temporal and spatial prediction ability, and long-term prediction ability.

Table I shows that the result of the five metrics using to measure the prediction result of the STAGCN model compared with other models on the East China Sea dataset for the 1-day, 7-days, 14-days, and 30-days prediction tasks, with boldface sections representing the optimal values of the various metrics of the model. - show that the value is negative, indicating that the model cannot predict well and can be ignored. Compared with other models, the proposed STAGCN model achieves excellent prediction performance under almost all conditions, indicating that the STAGCN model can capture the global spatio-temporal correlation, thus achieved an accurate prediction of SST. For different prediction lengths, the prediction effects of the STAGCN model are preferable to other models, demonstrating that the STAGCN model has the ability to predict SST accurately in both the short and long term.

As can be seen from Table I, the prediction performance of the GRU model is better than that of the GCN model. The RMSE of the GRU model is approximately 0.07 lower than that of the GCN model and the accuracy of the GRU model is increased by 0.4% compare with the GCN model for future one-day SST prediction. The forecasted effect of the GCN model is worse than that of the GRU model probably because the data itself has obvious time series features, but the GCN model merely captures the spatial dependency without considering the temporal characteristics from SST data.

TABLE I. THE PREDICTION RESULT OF THE STAGCN MODEL COMPARED WITH OTHER MODELS

Time	metrics	ARIMA	SVR	GCN	GRU	LSTM	STAGCN
1 day	RMSE	7.2835	0.9280	0.5034	0.4217	0.5528	0.3479
	MAE	5.9413	0.7567	0.3574	0.2825	0.4316	0.2354
	Accuracy	0.6625	0.9588	0.9759	0.9799	0.9757	0.9846
	R^2	-	0.9670	0.9933	0.9953	0.9884	0.9954
	Var	-	0.9671	0.9933	0.9953	0.9910	0.9954
7 day	RMSE	7.9466	1.2042	0.8746	0.8034	0.7909	0.6644
	MAE	6.5981	0.9860	0.6450	0.5836	0.6077	0.4907
	Accuracy	0.6304	0.9464	0.9581	0.9615	0.9651	0.9705
	R^2	-	0.9447	0.9798	0.9830	0.9761	0.9832
	Var	0.0090	0.9477	0.9799	0.9831	0.9786	0.9832
14 day	RMSE	7.7821	1.3285	1.0467	0.9493	0.8941	0.8122
	MAE	6.4279	1.0855	0.7860	0.7086	0.6853	0.6142
	Accuracy	0.6386	0.9408	0.9497	0.9544	0.9604	0.9640
	R^2	-	0.9325	0.9711	0.9762	0.9691	0.9749
	Var	-	0.9351	0.9711	0.9762	0.9691	0.9749
30 day	RMSE	8.3632	1.5406	1.3253	1.1428	1.1605	0.9694
	MAE	7.0629	1.2533	1.0052	0.8662	0.8860	0.7510
	Accuracy	0.6128	0.9314	0.9364	0.9452	0.9492	0.9568
	R^2	-	0.9095	0.9538	0.9657	0.9469	0.9642
	Var	0.0040	0.9117	0.9539	0.9666	0.9471	0.9642

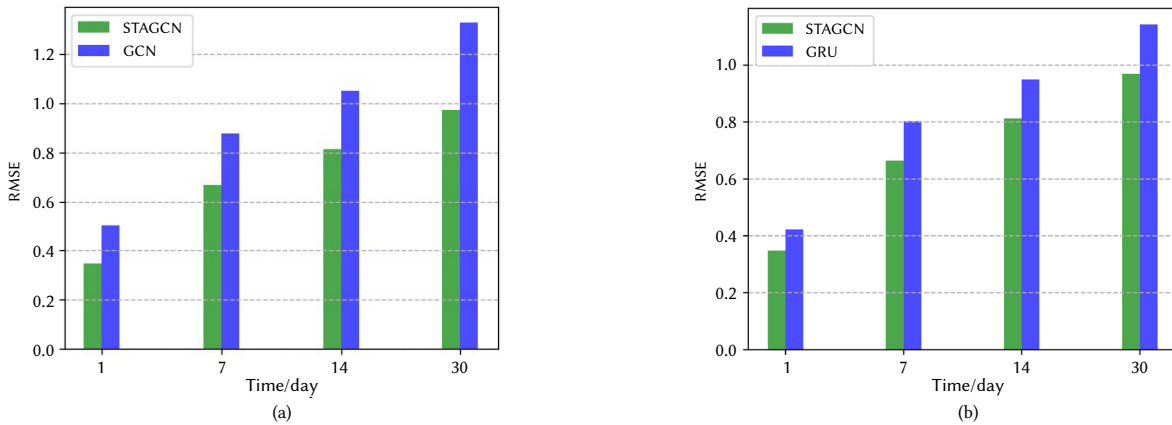


Fig. 5. Comparison of SST prediction errors between STAGCN model, GCN model and GRU model. (a) Comparison of RMSE values between STAGCN model and GCN model. (b) Comparison of RMSE values between STAGCN model and GRU model.

In order to better illustrate that the STAGCN model proposed in this paper can capture the temporal and spatial dependence from SST data and obtain satisfactory prediction effect simultaneously, we visualized the error index RMSE of STAGCN model, GRU model, and GCN model in predicting SST for next 1 day, 7 days, 14 days and 30 days and analyzed their prediction performance. The visualization results are shown in Fig. 5. Fig. 5 (a) and Fig. 5 (b) are visual comparison effects of the STAGCN model compared with the GCN model and the GRU model on RMSE metric respectively. As can be seen from the bar chart, the RMSE of the models increases with the increase of the predicted length, but the error of the STAGCN model is smaller than that of the other two models demonstrating that the STAGCN model has the ability to obtain spatio-temporal correlation from SST series data. For example, the RMSE values of STAGCN model are about 0.16, 0.21, 0.23 and 0.35 lower than that of the GCN model for 1-day, 7-day, 14-day and 30-day SST forecasting, indicating that the STAGCN model is capable to obtain spatial features from sequence data. The RMSE values of the GRU model considered single temporal characteristics are raised by about 0.07, 0.14, 0.14 and 0.17 for future 1-day, 7-day, 14-day and 30-

day SST prediction respectively compared with the STAGCN model, indicated that the STAGCN model can obtain the time correlation.

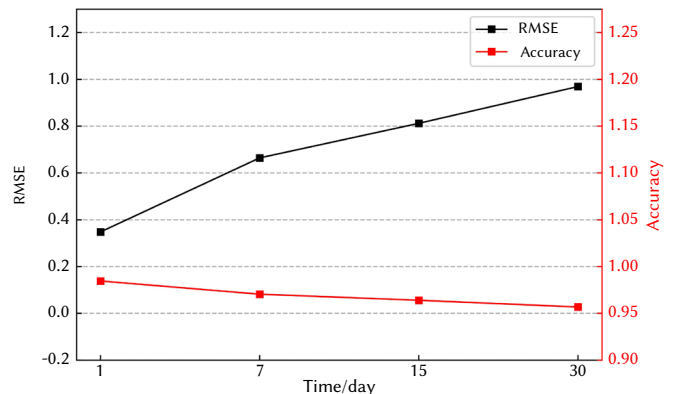


Fig. 6. The change of RMSE and Accuracy of STAGCN model for future SST prediction under different prediction interval conditions.

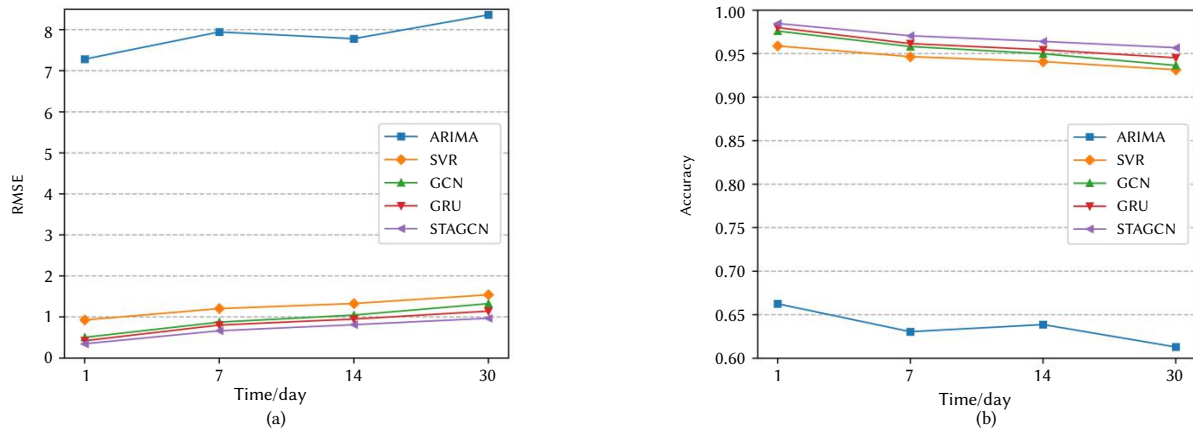


Fig. 7. The prediction performance of the STAGCN model and other methods under different prediction interval conditions.(a) Comparison of RMSE values between the STAGCN model and other methods. (b) Comparison of Accuracy between the STAGCN model and other methods.

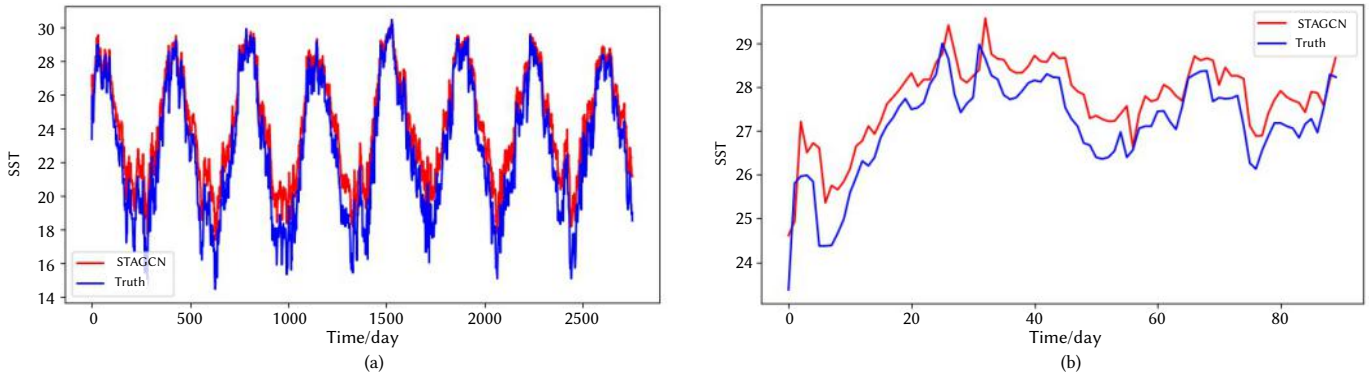


Fig. 8. The visualization comparison of the predicted and actual SST values for the next day.

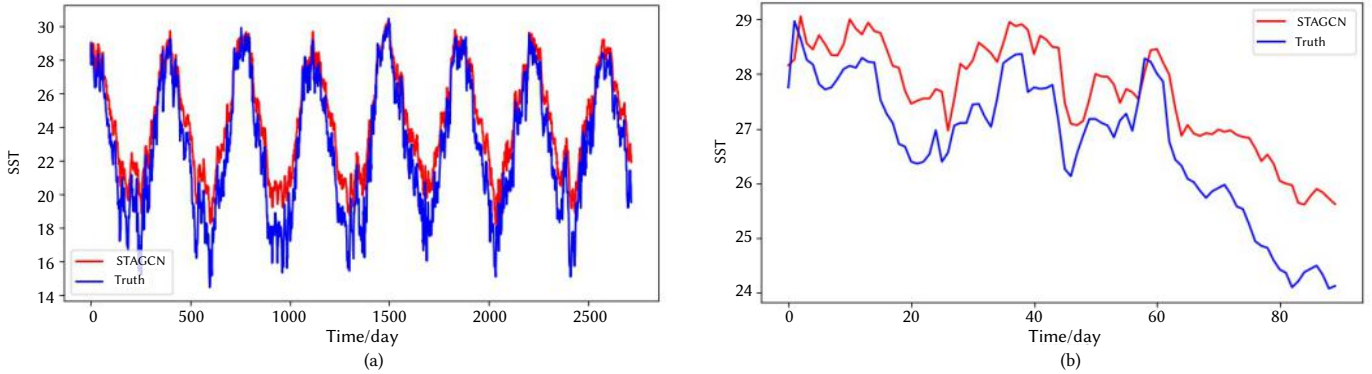


Fig. 9. The visualization comparison of the predicted and actual SST values for the 7 day.

To reflect the prediction effect of the STAGCN model, we visualized the variation trend of the RMSE and the accuracy in SST forecasting for the next 1 day, 7 days, 14 days and 30days. The results are shown in Fig.6. As can be seen from the figure, with the increase of prediction length, the error of the STAGCN model increases relatively large and the accuracy decreases slightly. Although the RMSE of the STAGCN model does not have a certain stability, its prediction accuracy is relatively stable, indicating that the STAGCN model can also achieve long-term SST prediction while it has better prediction ability for short-term SST prediction than long-term SST prediction.

Fig.7 (a) and Fig.7 (b) are RMSE and Accuracy results of SST prediction by different methods at different prediction horizons. It's observed that the STAGCN model achieved the lowest RMSE and the highest Accuracy for different predicted lengths.

To better explain the prediction performance of the STAGCN model, an ocean location point was randomly selected from the East

China Sea dataset, in which we predicted the future SST for the next 1 day, 7 days, 14 days, and 30 days, and visualize the prediction effect of all the selected days and the next 90 days. Fig. 8 , Fig. 9 , Fig. 10 , and Fig. 11 show the visualization results of the SST for the 1-day, 7-days, 14-days and 30-days forecast intervals.

The prediction results of STAGCN model with the prediction length of 1 day, 7 days, 14 days and 30 days show that STAGCN model has poor prediction at peak and peak valley. The main reason may be that the GCN model in STAGCN model captures spatial features by constantly moving its defined smoothing filter, which will lead to excessive peak smoothing of the overall prediction results. At the same time, the model has a corresponding delay in the overall prediction. Although the prediction performance of STAGCN model decreases with the increase of prediction length, the fitting degree of predicted value and real value of STAGCN model is still high, and good prediction results can be obtained.

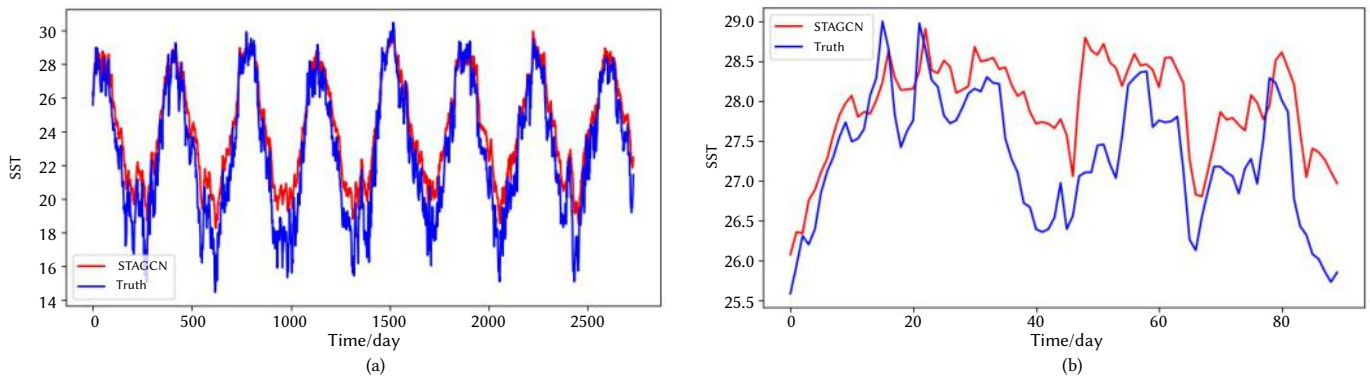


Fig. 10. The visualization comparison of the predicted and actual SST values for the 14 day.

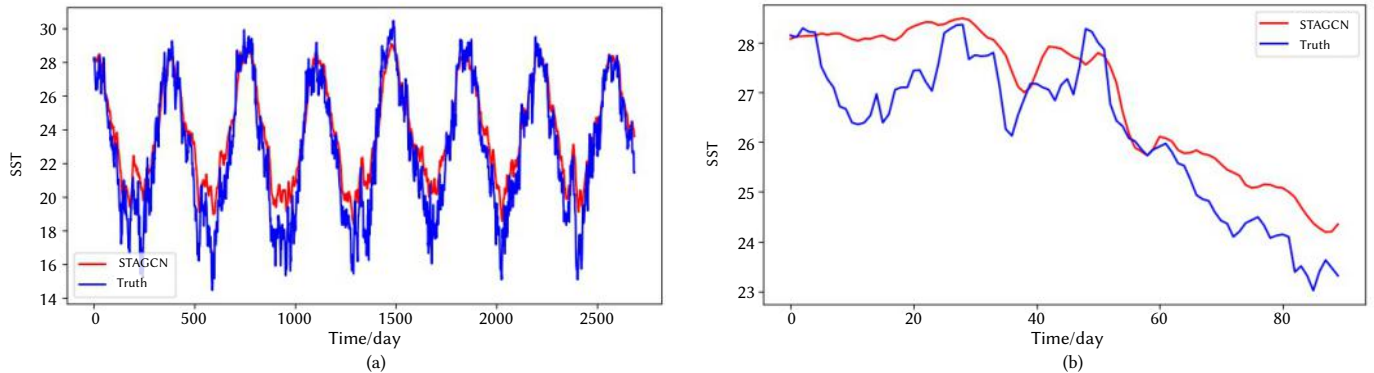


Fig. 11. The visualization comparison of the predicted and actual SST values for the 30 day.

STAGCN model can capture the temporal correlation and regional spatial characteristics from the TIME series of SST, and obtain the global spatiotemporal dynamic change trend by using the attention mechanism, which reduces the prediction error and improves the accuracy of prediction, and realizes the long-term and short-term prediction task of regional scale SST.

IV. CONCLUSION

The sea surface temperature is an important index to detect ocean changes, predict SST anomalies, and prevent natural disasters caused by abnormal changes, the dynamic variation of which have a profound impact on the whole marine ecosystem and the changes of climate. Therefore, it's essential to predict the future sea surface temperature. In order to achieve accurate SST prediction, a prediction model combining the GCN model with the GRU model and introduces the attention mechanism named the STAGCN model is proposed in this paper. We use the graph network to model the network of ocean location points. Nodes on the graph represent each ocean location point, edges on the graph represent that there have connections between location points. The GCN model is used to obtain the spatial correlation from the SST time series by constructing the spatial topology structure on the ocean points graph, which is obtained by the distance function between the nodes. The STAGCN model takes the GRU model to capture time dependence in the way of filtering and retaining historical and current SST information. Meanwhile, the attention model is applied to captures the importance of SST information from the output state and combines the global spatio-temporal characteristics from SST information. In this study, the experimental results of predicting the future short-term and long-term SST with STAGCN model on the data set indicating that the STAGCN model can achieve desirable prediction performance compared with the ARIMA model, SVR model, GCN model, and GRU model. In conclusion, the STAGCN model can acquire preferable forecasting results for future short-term and long-term SST

prediction in the way of capturing global spatial characteristics and temporal dependence from SST series data.

REFERENCES

- [1] A. J. Constable, Melbourne-Thomas, "Climate change and southern ocean ecosystems i: how changes in physical habitats directly affect marine biota," *Global change biology*, vol. 20, pp. 3004–3025, 1 2014.
- [2] M. Sumner, K. J. Michael, C. Bradshaw, M. A. Hindell, "Remote sensing of southern ocean sea surface temperature: implications for marine biophysical models," *Remote Sensing of Environment*, vol. 84, no. 2, pp. 161–173, 2003.
- [3] J. Yang, J. Wen, Y. Wang, B. Jiang, H. Wang, H. Song, "Fog-based marine environmental information monitoring toward ocean of things," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4238–4247, 2020.
- [4] M. Bouali, O. T. Sato, P. S. Polito, "Temporal trends in sea surface temperature gradients in the south atlantic ocean," *Remote Sensing of Environment*, vol. 194, pp. 100–114, 2017.
- [5] M. A. Cane, A. Kaplan, D. Pozdnyakov, S. E. Zebiak, "Twentieth-century sea surface temperature trends," *Science*, vol. 275, no. 5302, pp. 957–960, 1997.
- [6] G. A. Meehl, "Development of global coupled ocean- atmosphere general circulation models," *Climate Dynamics*, vol. 5, no. 1, pp. 19–33, 1990.
- [7] S. L. Castro, G. A. Wick, M. Steele, "Validation of satellite sea surface temperature analyses in the beaufort sea using uptoempo buoys," *Remote Sensing of Environment*, vol. 187, pp. 458–475, 2016.
- [8] R. Salles, P. Mattos, A. Iorgulescu, E. Bezerra, L. Lima, E. Ogasawara, "Evaluating temporal aggregation for predicting the sea surface temperature of the atlantic ocean," *Ecological Informatics*, vol. 36, pp. 94–105, 2016.
- [9] C. Xiao, N. Chen, C. Hu, K. Wang, Z. Chen, "Short and mid-term sea surface temperature prediction using time-series satellite data and lstm-adaboost combination approach," *Remote Sensing of Environment*, vol. 233, p. 111358, 2019.
- [10] N. Chen, K. Wang, C. Xiao, J. Gong, "A heterogeneous sensor web node meta-model for the management of a flood monitoring system," *Environmental Modelling Software*, vol. 54, no. apr., pp. 222–237, 2014.

- [11] K. Patil, M. C. Deo, M. Ravichandran, "Prediction of sea surface temperature by combining numerical and neural techniques," *Journal of Atmospheric and Oceanic Technology*, vol. 33, no. 8, pp. 1715–1726, 2016.
- [12] Y. Xue, A. Leetmaa, "Forecasts of tropical pacific sst and sea level using a markov model," *Geophysical Research Letters*, vol. 27, no. 17, pp. 2701–2704, 2000.
- [13] X. Yan, A. Leetmaa, J. Ming, "2000: Enso prediction with markov models: The impact of sea level," 2013, doi: [https://doi.org/10.1175/1520-0442\(2000\)013%3C0849:EPWMMT%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013%3C0849:EPWMMT%3E2.0.CO;2).
- [14] D. C. Collins, C. J. C. Reason, F. Tangang, "Predictability of indian ocean sea surface temperature using canonical correlation analysis," *Climate Dynamics*, vol. 22, no. 5, pp. 481–497, 2004.
- [15] T. Laepple, S. Jewson, "Five year ahead prediction of sea surface temperature in the tropical atlantic: a comparison between ipcc climate models and simple statistical methods," *Physics*, 2007, doi: <https://doi.org/10.48550/arXiv.physics/0701165>.
- [16] K. Patil, M. C. Deo, S. Ghosh, M. Ravichandran, "Predicting sea surface temperatures in the north indian ocean with nonlinear autoregressive neural networks," *International Journal of Oceanography*, 2013, (2013-4-30), vol. 2013, pp. 1–11, 2013.
- [17] I. D. Lins, M. Araujo, M. Moura, M. A. Silva, E. L. Drogue, "Prediction of sea surface temperature in the tropical atlantic by support vector machines," *Computational Statistics Data Analysis*, vol. 61, pp. 187–198, 2013.
- [18] A. Wu, W. W. Hsieh, B. Tang, "Neural network forecasts of the tropical pacific sea surface temperatures," *Neural Networks*, vol. 19, no. 2, pp. 145–154, 2006.
- [19] S. G. Aparna, S. D'Souza, N. B. Arjun, "Prediction of daily sea surface temperature using artificial neural networks," *International Journal of Remote Sensing*, vol. 39, no. 11-12, pp. 4214–4231, 2018.
- [20] Q. He, C. Zha, W. Song, Z. Hao, Y. Du, A. Liotta, C. Perra, "Improved particle swarm optimization for sea surface temperature prediction," *Energies*, vol. 13, 2020.
- [21] J. Xie, J. Zhang, J. Yu, L. Xu, "An adaptive scale sea surface temperature predicting method based on deep learning with attention mechanism," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 5, pp. 740–744, 2020.
- [22] X. Liu, T. Wilson, P. N. Tan, L. Luo, "Hierarchical lstm framework for long-term sea surface temperature forecasting," 2020.
- [23] J. Liu, T. Zhang, G. Han, Y. Gou, "Tlstm: Temporal dependence-based lstm networks for marine temperature prediction," *Sensors*, vol. 18, 2018.
- [24] Park, Kim, Lee, Song, "Temperature prediction using the missing data refinement model based on a long short-term memory neural network," *Atmosphere*, vol. 10, no. 11, p. 718, 2019.
- [25] S. Hochreiter, J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] R. Dey, F. M. Salemt, "Gate-variants of gated recurrent unit (gru) neural networks," pp. 1597–1600, 2017.
- [27] Z. Qin, W. Hui, J. Dong, G. Zhong, S. Xin, "Prediction of sea surface temperature using long short-term memory," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1745–1749, 2017.
- [28] Y. Feng, T. Sun, J. Dong, C. Li, "Study on long term sea surface temperature (sst) prediction based on temporal convolutional network (tcn) method," *ACM Turing Award Celebration Conference-China (ACM TURC 2021)*, pp. 28–32, 2021.
- [29] M. Han, Y. Feng, X. Zhao, C. Sun, C. Liu, "A convolutional neural network using surface data to predict subsurface temperatures in the pacific ocean," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2019.
- [30] M. Abadi, P. Barham, J. Chen, Z. Chen, X. Zhang, "Tensorflow: A system for large-scale machine learning," 2016.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, "Scikit-learn: Machine learning in python," 2012.



Jiabao Wen

He received the Ph.D. degree in information and communication engineering from Tianjin University, China, in 2021. He is currently a postdoctor at School of Electrical and Information Engineering, Tianjin University. His research interests include ocean information processing, pattern recognition, cloud computing.



Caiyun Lv

She received the Master degree in Electronics and Communication Engineering from Tianjin University, China, in 2022. Her research interests include ocean prediction and ocean information processing.



Desheng Chen

He is currently pursuing the Ph.D. degree at the School of Electrical and Information Engineering, Tianjin University, China. His research interests include marine exploration, marine intelligent equipment control, artificial intelligence and Big data computing.