Sea Surface Temperature Prediction With Memory Graph Convolutional Networks

Xiaoyu Zhang, Yongqing Li, Alejandro C. Frery, *Senior Member, IEEE*, and Peng Ren, *Senior Member, IEEE*

***Abstract*— We develop a memory graph convolutional network (MGCN) framework for sea surface temperature (SST) prediction. The MGCN consists of two memory layers: one graph layer and one output layer. The memory layer captures SST temporal changes via temporal convolution units and gate linear units. The graph layer encodes SST spatial changes in terms of characteristics derived from graph Laplacian. The output layer encapsulates information from the previous layers and produces SST prediction results. The MGCN characterizes both the temporal and spatial changes, rendering a comprehensive SST prediction strategy. We use daily mean SST data for two areas near the Bohai Sea and the East China Sea for experimental evaluations and validate that the MGCN performs better than other traditional machine learning methods for nearshore SST prediction. In addition, we test the MGCN on weekly and monthly mean SST datasets and validate that the MGCN is robust and suitable for SST prediction.**

***Index Terms*— Graph convolutional network, sea surface temperature (SST), spatiotemporal prediction.**

I. Introduction

SEA surface temperature (SST) is an essential variable for global climate and weather changes studies. Predicting SST is an important topic of ocean research [1]. There are two major goals of developing SST prediction methods. The first is to promote the understanding of ocean dynamics; this requires characterizing large-scale, high-resolution SST variations. The second goal, which is our focus, is to provide effective tools to marine applications. For instance, SST predictions can provide meaningful indications for fishing directions [2] and storm tracking [3]. In this scenario, the scale and resolution required for SST prediction are flexible, as they are usually employed as a layer of information along with other variables as, for instance, chlorophyll concentration and atmospheric pressure.

SST prediction methods belong to two main categories: numerical and data-driven. Numerical methods utilize mathematical models that describe ocean dynamics and thermodynamics to establish prediction models [4]. Recently, Wang *et al.* [5] proposed a model-based analog forecasting (MAF) method to predict Indo-Pacific SST values. The model uses subsurface thermal conditions as one of the influencing factors for SST prediction. Numerical methods use intricate equations that tend to underestimate the observed SST variability. Data-driven methods learn the latent characteristics from SST data. Typically, Markov models [6] are constructed in terms of seasonally changing multivariable forecast. Multilayer perceptron regression (MLPR) [7] and support vector regression (SVR) [8] are widely used for SST prediction. Recurrent neural networks (RNNs) and their variants (long- and short-term memory—LSTM and gated recurrent units—GRUs), developed for learning time-series data changes. Zhang *et al.* [9] employed a fully connected LSTM to predict future SST values. The LSTM structure models SST sequences, and the fully connected structure produces the predictions. The GRU model is effective and has a simpler structure than the LSTM model. Zhang *et al.* [10] designed a neural network based on the GRU model for middle- and long-term SST prediction.

The aforementioned methods predict SST values without considering spatial factors. However, SST values of different sites are likely to be related [11], [12]. Xiao *et al.* [13] developed a convolutional LSTM for predicting the areal SST field. The model uses convolutional neural networks to extract the location information of the area on the basis of LSTM.

There are no valid SST data in some areas, such as land or islands. In this scenario, it is difficult for the convolutional neural network to fully encode SST spatial changes. In our work, we develop a memory graph convolutional (MGCN) network for SST prediction. There are two main contributions. First, we develop a model based on memory and graph layers for spatiotemporal SST prediction. The memory layer captures the temporal changes of the SST sequence, whereas the graph layers encode SST spatial changes in the frequency domain. Specifically, graph representations can be constructed in arbitrary forms that overcome the data irregularity. Second, we use a 39-year SST time series in the Bohai Sea and the East China Sea to validate the effectiveness of the proposed MGCN network for coastal SST prediction.

Manuscript received January 18, 2021; revised April 1, 2021 and May 27, 2021; accepted July 8, 2021. Date of publication July 26, 2021; date of current version January 6, 2022. This work was supported in part by the National Natural Science Foundation of China under Contract 61971444 and in part by the Shandong Provincial Natural Science Foundation under Grant ZR2019MF019. *(Corresponding author: Peng Ren.)*

Xiaoyu Zhang and Peng Ren are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China (e-mail: zhangxiaoyuupc@163.com; pengren@upc.edu.cn).

Yongqing Li is with the College of Control Science and Engineering, China University of Petroleum (East China), Qingdao 266580, China (e-mail: lyq_upc@163.com).

Alejandro C. Frery is with the School of Mathematics and Statistics, Victoria University of Wellington, Wellington 6140, New Zealand, and also with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xian 710126, China (e-mail: alejandro.frery@vuw.ac.nz).

Digital Object Identifier 10.1109/LGRS.2021.3097329

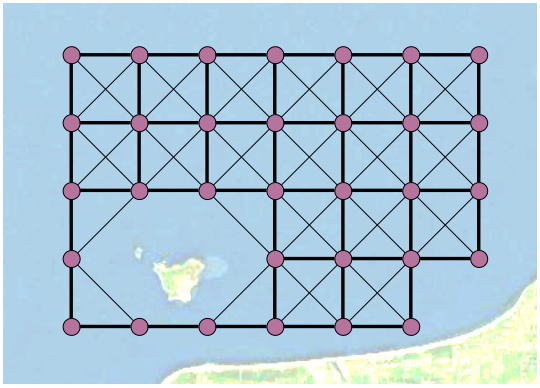

Fig. 1. Graph representation of SST data at a single instant.

Experimental results show that our model has a better performance than some commonly used machine learning methods.

II. Memory Graph Convolutional Networks for SST Prediction

A. Graph Representations for SST Data

We employ graphs for representing SST data. The sea surface locations are represented in terms of longitude and latitude grids. An SST value at each grid point for each time is recorded. In our work, we define graph representations in terms of valid SST data on grids.

Assume that we have a sequence $\boldsymbol{S} = (S_1, S_2, \ldots, S_T)$ of data defined on a regular grid of longitude and latitude coordinates, and each layer S_t is comprised of N valid SST points (locations, such as coasts and islands, do not render valid SST measurements). We define the adjacency matrix $\boldsymbol{W} = \left[w_{i,j}\right]$ for S_t, $1 \leq t \leq T$, based on the Euclidean distance $d_{i,j}$ between coordinates i and j as

$$w_{ij} = \begin{cases} \exp\left\{-\dfrac{d_{ij}^2}{r^2}\right\}, & \text{if } d_{ij} \leq d_{\min} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where r is a scaling factor and $d_{\min}$ is a threshold distance above which the weight is zero.

Fig. 1 shows this configuration for $r = 1$. The purple bullets denote the $N = 32$ grid valid points, for which $w_{i,j} = 1$. The thick segments denote points at distance $d_{1,2} = 1$, for which $w_{i,j} = \exp\{-1\}$. The thin segments denote points at distance $d_{1,2} = \sqrt{2}$, for which $w_{i,j} = \exp\{-2\}$. Pairs of points at distances greater than $d_{\min} = \sqrt{2}$ are not connected and, thus, their weights are zero.

An advantage of graph representations is that they comprehensively characterize irregular data. As illustrated, locations, such as coasts and islands without valid SST data, are not considered in the graph representation. In contrast to the full grid data requirement arising in [13], the graph representation is more flexible.

B. Memory Graph Convolutional Networks

Fig. 2 shows the main structure of the memory graph convolutional network (MGCN) for SST prediction. The input of the network is a sequence of T SST images $(S_1, S_2, \ldots, S_T)$ with adjacency matrix $\boldsymbol{W}$, computed according to (1). The output of the network is the SST prediction result S_{T+1} at time $T + 1$.

The network consists of two memory layers: one graph layer and one output layer. The memory and the graph layers capture the temporal and spatial changes of the SST sequence, respectively. The output layer maps the final prediction results. The graph layer is sandwiched between two memory layers. Such a structure realizes the rapid propagation of networks by scale compression [14].

Denote $\boldsymbol{S}^{(\ell)} \in \mathbb{R}^{N \times T_\ell \times C_\ell}$ as the input for each layer of the network, where N is the number of valid SST points and T_ℓ and C_ℓ denote the input length and the number of input channels at the ℓth layer, respectively.

C. Memory Layers for Capturing SST Temporal Changes

We employ memory layers to capture the temporal changes of the SST data sequence. The two memory layers have the same structure, shown in Fig. 3: a temporal convolution unit and a gated linear unit (GLU) [15].

The temporal convolution unit takes $\mathcal{S}^{(1)} = (S_1, S_2, \ldots, S_T)$ as input and produces the output $\boldsymbol{U}$ as follows:

$$\mathcal{U} = \boldsymbol{S}^{(1)} * \mathcal{A}_{\mathrm{m}} + \boldsymbol{b}_{\mathrm{m}} \tag{2}$$

where $\boldsymbol{A}_{\mathrm{m}}$ and $\boldsymbol{b}_{\mathrm{m}}$ denote the convolution kernels and the bias of the memory layer, respectively, and $*$ is the convolution operation with kernel of width K_{m} and height is set to 1. The output $\mathcal{U}$ is divided equally in channel dimension into two parts, $\mathcal{X}$ and $\mathcal{Y}$, which have the same number of channels. In addition, a K_{m}-length sequence $\mathcal{S}_{\mathrm{c}}^{(1)} = (S_{T-K_{\mathrm{m}}+1}, \ldots, S_T)$ is cropped from $\boldsymbol{S}^{(1)}$ such that $\mathcal{S}_{\mathrm{c}}^{(1)}$ and $\mathcal{X}$ (and $\mathcal{Y}$) have the same size. The GLU produces $\mathcal{S}^{(2)}$, the output of the first memory layer, as

$$\mathcal{S}^{(2)} = \left(\mathcal{X} + \mathcal{S}_{\mathrm{c}}^{(1)}\right) \odot \sigma(\mathcal{Y}) \tag{3}$$

where $\odot$ is the Hadamard product and σ is the sigmoid function.

In contrast to recurrent networks, memory layers only involve convolution without recurrence, rendering improved computational efficiency.

D. Graph Layers for Encoding SST Spatial Changes

We employ a graph layer to capture the spatial changes of the SST data sequence. Fig. 4 shows its structure.

We use spectrum graph convolution to encode SST spatial changes. The graph Laplacian matrix $\boldsymbol{L}$ for characterizing the graph spectrum is defined as follows:

$$\boldsymbol{L} = \boldsymbol{I}_N - \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{W} \boldsymbol{D}^{\frac{1}{2}} \tag{4}$$

where $\boldsymbol{I}_N$ is the identity matrix and $\boldsymbol{D}_{ii} = \sum_j w_{ij}$. The spectrum graph convolution operates in terms of a product of the weighted spatial convolution operator g_θ and the output of the memory layer $\mathcal{S}^{(2)}$

$$\mathcal{V} = g_\theta(\boldsymbol{L}) \mathcal{S}^{(2)} \tag{5}$$

where $\mathcal{V}$ denotes the output of the graph convolution. We use Chebyshev polynomials [16] to approximate the weighted

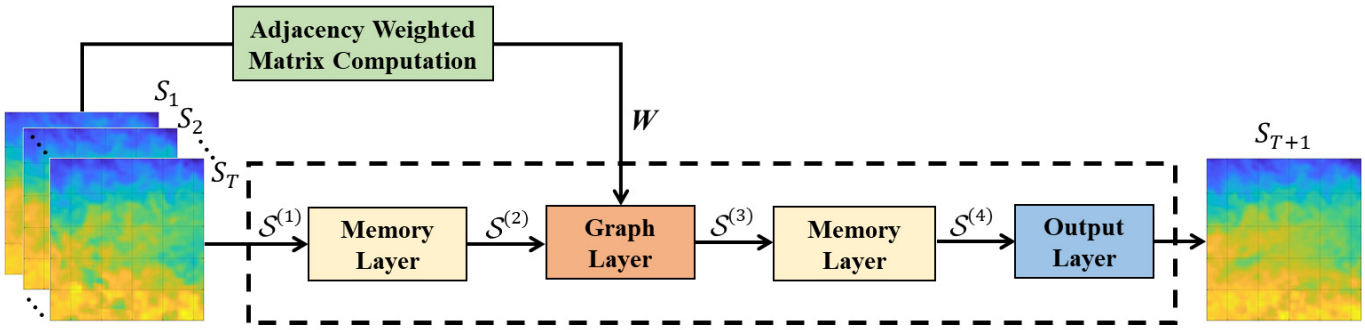

Fig. 2. MGCN (surrounded by the dashed box).

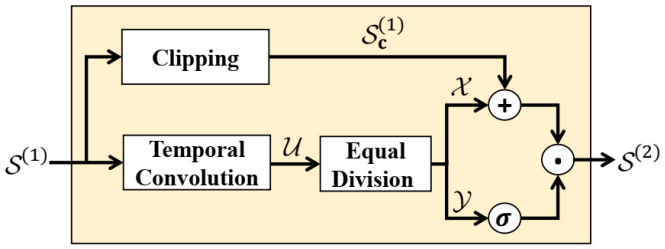

Fig. 3. Memory layer.

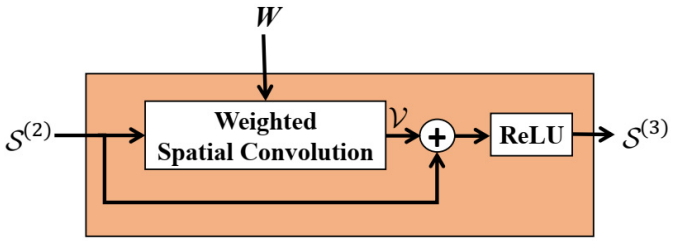

Fig. 4. Graph layer.

spatial convolution operator g_θ to reduce the computational complexity. The graph convolution unit is given as

$$\mathcal{V} = \sum_{k=1}^{K} T_k(2\boldsymbol{L}/\lambda_{\max} - \boldsymbol{I}_N)\mathcal{S}^{(2)}\theta_k \tag{6}$$

where $T_k(\cdot)$ denotes the Chebyshev polynomial of order k, θ_k denotes the matrix of polynomial coefficients of order k, and $\lambda_{\max}$ denotes the largest eigenvalue of $\boldsymbol{L}$.

We use the residual connection to improve network training accuracy. Overall, the graph convolution operates as follows:

$$\mathcal{S}^{(3)} = \mathrm{ReLu}(\mathcal{S}^{(2)} + \mathcal{V}) \tag{7}$$

where rectified linear unit (ReLU) denotes a nonlinear activation function.

In contrast to 2-D convolution methods that just conduct convolutions over regular sizes, the graph layer easily processes SST data at irregular locations and encode spatial dependencies among neighboring SST values in a comprehensive and controlled manner.

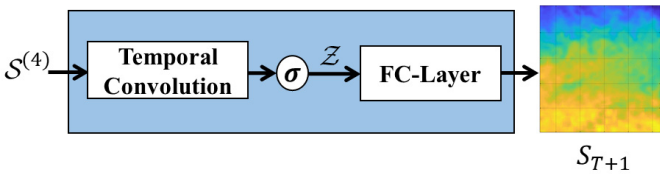

Fig. 5. Output layer.

E. Output Layers and Loss Functions

We construct the output layer based on a temporal convolutional unit with a sigmoid function and a fully connected layer, as shown in Fig. 5.

The convolution unit with a sigmoid function maps the output of the second memory layer $\mathcal{S}^{(4)}$ to a single-step prediction $\mathcal{Z}$. The fully connected layer produces the SST prediction $\widehat{S}_{T+1}$ as follows:

$$\widehat{S}_{T+1} = \mathcal{A}_o\mathcal{Z} + \boldsymbol{b}_o \tag{8}$$

where $\mathcal{A}_o$ and $\boldsymbol{b}_o$ denote the weight and bias of the output layer, respectively.

The loss function for training is given as

$$L(\Theta) = \|\mathrm{MGCN}((S_1, S_2, \ldots, S_T), \boldsymbol{W}; \Theta) - S_{T+1}\| + L_{\mathrm{reg}} \tag{9}$$

where Θ are trainable parameters and L_{reg} is a regularization term. We use L_2 regularization to avoid overfitting.

III. Experiments

A. Datasets and Experimental Settings

We use optimum interpolation SST (OISST) data produced by the National Oceanic and Atmospheric Administration (NOAA), Boulder, CO, USA (https://www.ncdc.noaa.gov/oisst). The NOAA OISST datasets contain daily, weekly, and monthly mean SST grid data. Daily OISST data covers the global ocean from 89.75°S to 89.75°N and 0.25°E to 359.25°E, with spatial resolution of 0.25° × 0.25°. Weekly and monthly OISST data cover the global ocean from 89.5°S to 89.5°N and 0.5°E to 359.5°E with spatial resolution of 1° × 1°.

We create SST datasets that contain SST sequences in the Bohai Sea and the East China Sea from NOAA OISST data. The temporal coverage of the SST datasets is from January 1982 to December 2019. The data for each region

include daily, weekly, and monthly mean SST datasets. Each SST dataset is further divided into three sets: training, validation, and test. The training sets are the SST data from 1982 to 2017, and the validation and test sets are the SST data for the entire year of 2018 and 2019, respectively. The Bohai Sea daily mean SST dataset covers the area from 37.25°N to 41.00°N and 117.50°E to 121.50°E. The Bohai Sea weekly and monthly mean SST datasets cover the area from 37°N to 41°N and 117°E to 121°E. The East China Sea daily mean SST dataset covers the area from 27.75°N to 32.50°N and 123.00°E to 126.75°E. The East China Sea weekly and monthly mean SST datasets cover the area from 28°N to 32°N and 123°E to 127°E. SST data are missing on some grid points (especially for lands or islands) in the experimental area. These points are not considered for computing the weighted adjacency matrix.

MGCN is compared with SVR, gate recurrent unit (GRU), and long and short-term time-series network (LSTNet) [17]. SVR and GRU are time-series predicting methods. LSTNet is a spatiotemporal predicting method that operates by extracting the local dependencies between SST. All methods are used to predict the SST of the next day, next three days, and next seven days based on the daily mean SST dataset. In addition, experiments are conducted for predicting SST for the next week and the next three weeks based on the weekly mean SST dataset and the next month based on the monthly mean SST dataset. All methods use manual search to set hyperparameters for getting the best performance.

We use the root-mean-squared error (RMSE) and the mean absolute error (MAE) to evaluate the effectiveness of the four methods. The smaller RMSE and MAE are, the better is method's performance.

We conducted the experiments on a computing platform with three Intel Xeon CPUs E5-2690 at 2.60 GHz and NVIDIA Tesla K80 GPU. GRU, LSTNet, and MGCN are implemented by Tensorflow under Python. SVR is implemented by Scikit-learn, a machine learning library based on Python. To support the reproducibility and replicability in remote sensing research [18], we have publicly released our code (https://github.com/upczxy/MGCN-SST-Prediction).

B. Results and Analysis

The setting of the adjacency matrix is the key to the MGCN network. We carry out experiments with different values of r and $d_{\min}$ in (1) to obtain the best network performance. We calculate the RMSE of the predicted SST from the Bohai Sea and the East China Sea datasets for the next seven days.

Denote the threshold $\varphi = \exp\{-d_{\min}^2/r^2\}$. We set $r = 1, 5, 10$ according to the area size, and φ is set from 0 to 0.9. Fig. 6 shows the RMSE of the seven day predictions with varying r and φ. The results show that $(r = 1, \varphi = 0.7)$ and $(r = 5, \varphi = 0.2)$ are suitable parameters for the Bohai Sea SST and the East China Sea SST predictions, respectively.

Table I shows the performance of the next day, the next three days, and the next seven days' predictions with the four methods on the Bohai Sea and East China Sea SST datasets. The bold entries denote the best results. It is clear that the prediction performance of our MGCN method is better than

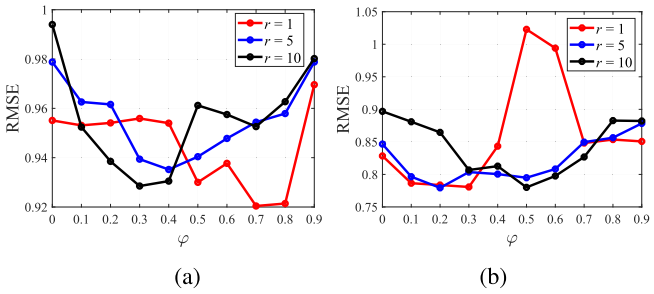

Fig. 6. RMSE of seven day predictions with different values of r and φ. (a) Bohai Sea SST. (b) China East Sea SST.

TABLE I
DAILY PREDICTION RESULTS (RMSE AND MAE) ON THE BOHAI SEA AND CHINA EAST SEA SST DATASETS

Sea Area	Method	Metrics	Daily		
			1	3	7
Bohai Sea	SVR	RMSE	0.5179	0.8183	1.2145
		MAE	0.4026	0.6147	0.9630
	GRU	RMSE	0.4111	0.7808	1.0681
		MAE	0.3021	0.5946	0.8420
	LSTNet	RMSE	0.3907	0.7413	1.0593
		MAE	0.2719	0.5589	0.8208
	MGCN	RMSE	**0.3858**	**0.6950**	**0.9204**
		MAE	**0.2711**	**0.5314**	**0.7206**
East China Sea	SVR	RMSE	0.3746	0.6601	0.9182
		MAE	0.2697	0.4940	0.7022
	GRU	RMSE	0.4689	0.6739	0.9557
		MAE	0.3854	0.5082	0.7202
	LSTNet	RMSE	0.4458	0.6785	0.9056
		MAE	0.3404	0.5195	0.7035
	MGCN	RMSE	**0.3274**	**0.6026**	**0.7637**
		MAE	**0.2229**	**0.4555**	**0.5941**

the other three methods over the three prediction time scales. Specifically, the other three methods have different prediction performances for the Bohai Sea and the East China Sea datasets. For example, the LSTNet method outperforms SVR on the Bohai Sea SST dataset, but the performance is opposite for one- and three-day predictions on the East Sea SST dataset. Our MGCN model performs best on both datasets compared with the other three methods. It reflects that MGCN is suitable for SST predictions in different areas.

Fig. 7 shows the RMSE for 1–7 days ahead SST prediction using four methods. The RMSE of the MGCN network increases at a smaller rate than those of the reference methods as the prediction horizon increases from 1 to 7 days. The figure shows that MGCN maintains a respectable performance for long-time daily mean SST predictions.

Table II presents the performance of the four methods on the prediction of SST datasets in one week, three weeks, and one month. MGCN outperforms the other three methods, and SVR is the worst one. Besides, the temporal and spatial resolution of weekly and monthly mean SST data is lower than daily mean SST data, and the amount of data is also reduced. The prediction accuracy of the reference methods decreases. It confirms that MGCN is more accurate and robust

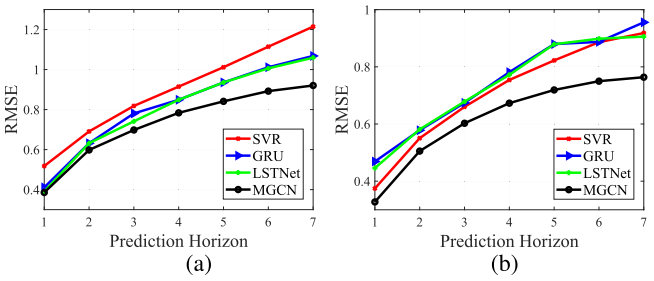

Fig. 7. RMSE for 1–7 days ahead SST prediction by four methods. (a) Bohai Sea SST. (b) China East Sea SST.

TABLE II
WEEKLY AND MONTHLY PREDICTION RESULTS (RMSE AND MAE) ON THE BOHAI SEA AND CHINA EAST SEA SST DATASETS

Sea Area	Method	Metric	Weekly		Monthly
			1	3	1
Bohai Sea	SVR	RMSE	1.1165	1.4248	1.0372
		MAE	0.9680	1.2065	0.8822
	GRU	RMSE	0.7540	0.9493	0.6581
		MAE	0.6400	0.7506	0.7526
	LSTNet	RMSE	0.5563	0.9200	0.5796
		MAE	0.4500	0.7178	0.4958
	MGCN	RMSE	**0.4699**	**0.7861**	**0.4994**
		MAE	**0.3717**	**0.6210**	**0.4528**
East China Sea	SVR	RMSE	0.9017	1.1664	1.0911
		MAE	0.7542	0.9509	0.8929
	GRU	RMSE	0.7632	1.0136	0.7778
		MAE	0.6115	0.8279	0.6395
	LSTNet	RMSE	0.6662	1.0035	0.7667
		MAE	0.5342	0.8341	0.6310
	MGCN	RMSE	**0.5588**	**0.7351**	**0.6365**
		MAE	**0.4399**	**0.6049**	**0.5333**

in long-term SST prediction. Furthermore, the method is better suited for monthly timescales than daily and weekly timescales because the latter are chaotic in nature.

The computational complexity for training the entire network is of order $O(ENTK_mC_3C_4)$, where E denotes the training iterations and C_3 and C_4 denote the number of input channels at the graph layer and the second memory layer, respectively. The adjacency weighted matrix is computed only once, with computational complexity $O(N^2)$. Compared with recurrent networks, our method achieves parallelization input and improves training efficiency.

IV. CONCLUSION

In this letter, an MGCN has been developed to predict future SST values by learning temporal changes with spatial information. The network consists of two memory layers, one graph layer, and one output layer. The memory layers are used to capture the time-series changes of SST, and the graph layer learns the spatial relationship of SST. The output layer maps the prediction result.

The experiments explore the optimal parameters of the adjacency matrix in the convolutional graph and validate the effectiveness of the MGCN on the SST data from the Bohai Sea and East China Sea. Experimental results confirm that our MGCN method improves the RMSE on different time scales compared with the other three methods.

The suitable adjacency matrix is different for different resolutions and areas of SST data. We will investigate more comprehensive ways for learning weights to improve the SST prediction effectiveness. Furthermore, low-level remote sensing SST observation products are disturbed by cloud coverage, and there might be blank in some regions. In future work, we will use a graph layer to extract spatial features of SST data. Then, we will feed the spatial features to a memory layer implementing a diffusion process that progressively reconstructs the score matrix. Finally, we will reconstruct and predict cloud coverage SST data.

REFERENCES

[1] C. Zhang, "Large-scale variability of atmospheric deep convection in relation to sea surface temperature in the tropics," *J. Climate*, vol. 6, no. 10, pp. 1898–1913, Oct. 1993.
[2] J. P. Eveson, A. J. Hobday, J. R. Hartog, C. M. Spillman, and K. M. Rough, "Seasonal forecasting of tuna habitat in the great australian bight," *Fisheries Res.*, vol. 170, pp. 39–49, Oct. 2015.
[3] A. Kuwano-Yoshida and S. Minobe, "Storm-track response to SST fronts in the northwestern pacific region in an AGCM," *J. Climate*, vol. 30, no. 3, pp. 1081–1102, Feb. 2017.
[4] W. Peng, Q. Chen, S. Zhou, and P. Huang, "CMIP6 model-based analog forecasting for the seasonal prediction of sea surface temperature in the offshore area of China," *Geosci. Lett.*, vol. 8, no. 1, pp. 1–8, Dec. 2021.
[5] Y. Wang, Z. Zhang, and P. Huang, "An improved model-based analogue forecasting for the prediction of the tropical Indo-Pacific sea surface temperature in a coupled climate model," *Int. J. Climatol.*, vol. 40, no. 15, pp. 6346–6360, Dec. 2020.
[6] Y. Xue and A. Leetmaa, "Forecasts of tropical pacific SST and sea level using a Markov model," *Geophys. Res. Lett.*, vol. 27, no. 17, pp. 2701–2704, Sep. 2000.
[7] S. Chaudhuri, S. Goswami, and A. Middey, "Medium-range forecast of cyclogenesis over North Indian Ocean with multilayer perceptron model using satellite data," *Natural Hazards*, vol. 70, no. 1, pp. 173–193, Jan. 2014.
[8] M. Imani *et al.*, "Spatiotemporal prediction of satellite altimetry sea level anomalies in the tropical Pacific ocean," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1126–1130, Jul. 2017.
[9] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1745–1749, Oct. 2017.
[10] Z. Zhang, X. Pan, T. Jiang, B. Sui, C. Liu, and W. Sun, "Monthly and quarterly sea surface temperature prediction based on gated recurrent unit neural network," *J. Mar. Sci. Eng.*, vol. 8, no. 4, pp. 249–263, 2020.
[11] G. Zheng, X. Li, R. H. Zhang, and B. Liu, "Purely satellite data–driven deep learning forecast of complicated tropical instability waves," *Sci. Adv.*, vol. 6, no. 29, p. 1482, 2020.
[12] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, "A CFCC-LSTM model for sea surface temperature prediction," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 207–211, Feb. 2018.
[13] C. Xiao *et al.*, "A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data," *Environ. Model. Softw.*, vol. 120, Oct. 2019, Art. no. 104502.
[14] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*. [Online]. Available: https://arxiv.org/abs/1709.04875
[15] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. 34th Annu. Int. Conf. Mach. Learn.*, 2017, pp. 933–941.
[16] W. Peng, X. Hong, H. Chen, and G. Zhao, "Learning graph convolutional network for skeleton-based human action recognition by neural searching," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 2669–2676.
[17] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jun. 2018, pp. 95–104.
[18] A. C. Frery, L. Gomez, and A. C. Medeiros, "A badging system for reproducibility and replicability in remote sensing research," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4988–4995, Aug. 2020.