Enhancing aquaculture water quality forecasting using novel adaptive multi-channel spatial-temporal graph convolutional network

Tianqi Xiang[1], Xiangyun Guo[2*], Junjie Chi[3], Juan Gao[1], Luwei Zhang[3]

(1. *College of Computer Science, Beijing Information Science & Technology University, Beijing 102206, China*;
2. *College of Management Science and Engineering, Beijing Information Science & Technology University, Beijing 102206, China*;
3. *College of Engineering, China Agricultural University, Beijing 100083, China*)

Abstract: In recent years, aquaculture has developed rapidly, especially in coastal and open ocean areas. In practice, water quality prediction is of critical importance. However, traditional water quality prediction models face limitations in handling complex spatiotemporal patterns. To address this challenge, a prediction model was proposed for water quality, namely an adaptive multi-channel temporal graph convolutional network (AMTGCN). The AMTGCN integrates adaptive graph construction, multi-channel spatiotemporal graph convolutional network, and fusion layers, and can comprehensively capture the spatial relationships and spatiotemporal patterns in aquaculture water quality data. Onsite aquaculture water quality data and the metrics MAE, RMSE, MAPE, and R^2 were collected to validate the AMTGCN. The results show that the AMTGCN presents an average improvement of 34.01%, 34.59%, 36.05%, and 17.71% compared to LSTM, respectively; an average improvement of 64.84%, 56.78%, 64.82%, and 153.16% compared to the STGCN, respectively; an average improvement of 55.25%, 48.67%, 57.01%, and 209.00% compared to GCN-LSTM, respectively; and an average improvement of 7.05%, 5.66%, 7.42%, and 2.47% compared to TCN, respectively. This indicates that the AMTGCN, integrating the innovative structure of adaptive graph construction and multi-channel spatiotemporal graph convolutional network, could provide an efficient solution for water quality prediction in aquaculture.

Keywords: water quality prediction, aquaculture, spatial-temporal graph convolutional network, multi-channel, adaptive graph construction

DOI: 10.25165/j.ijabe.20251801.9074

Citation: Xiang T Q, Guo X Y, Chi J J, Gao J, Zhang L W. Enhancing aquaculture water quality forecasting using novel adaptive multi-channel spatial-temporal graph convolutional network. Int J Agric & Biol Eng, 2025; 18(1): 279–291.

1 Introduction

The marine aquaculture industry is expanding to coastal and open ocean areas due to rising demand for seafood and restrictions on wild catch fisheries[1-3]. Large-scale deep-sea cages and engineered enclosures are crucial for farming fish in the open ocean due to the vast volume of water, expansive activity space for aquaculture organisms, efficient exchange of aquaculture water, high utilization rate of residual feed, and the full utilization of natural resources in the marine environment. These enclosures allow for the rapid growth of aquaculture animals. Simultaneously, organism quality is closer to the natural ecology[4]. All of these advantages make breeding organisms in the open ocean an important aspect in compensating for the depletion of marine catches[5].

In mariculture, seawater serves as the primary medium for the habitat and material exchange of aquaculture organisms, which are highly sensitive to variations in the physical and chemical factors of water. Even slight fluctuations outside the optimal conditions can cause physiological stress in the organisms, leading to reduced food intake, increased energy consumption, and susceptibility to diseases. For example, inadequately dissolved oxygen (DO) can cause slow fish growth and even mortality. Conversely, excessively high DO can make fish susceptible to bubble disease[6]. Large changes in DO, temperature, pH, and other water quality factors may directly lead to the death of aquaculture organisms. Meanwhile, changes in the water quality parameters may lead to the deterioration of the aquaculture ecological environment, such as outbreaks of red tide algae, bacteria, parasites, etc.[7,8] Meanwhile, water parameters in aquaculture are multi-correlated, highly complex, and unstable[9]. Increased pH and water temperature can lead to an increased ionization fraction of NH_3 in water[10,11]. The DO content is easily affected by other water quality factors such as pH and temperature, and it has the characteristics of nonlinearity, a large time lag, and instability[6]. Therefore, accurately predicting water quality data in marine ranches, together with considering the dynamic relationships between the water quality parameters, is a pressing issue in the aquaculture industry and could facilitate an early warning system, further providing a basis for aquaculture management decisions and realizing sustainable development[7,8,12].

Currently, there are three main categories of water quality prediction methods: traditional time series models, machine learning methods, and deep learning methods[13]. Traditional time series models primarily rely on autoregressive integrated moving average (ARIMA) models and their various variants[13]. For instance, Wang et al. applied an ARIMA model to dynamic changes in

---

Received date: 2024-05-14 **Accepted date:** 2024-12-17

Biographies: Tianqi Xiang, Master candidate, research interest: deep learning, information processing, Email: xiangtianqi2019@163.com; **Junjie Chi**, Master, research interest: smart agriculture system integration research, Email: chijunjie1018@163.com; **Juan Gao**, Master candidate, research interest: deep learning, information processing, Email: 1720824512@qq.com; **Luwei Zhang**, PhD candidate, research interest: information processing, sensors technology, Email: zhreed@126.com.

***Corresponding author: Xiangyun Guo**, Associate Professor, research interest: precision agriculture, information processing. College of Management Science and Engineering, Beijing Information Science and Technology University. Tel: +86-13426005302, Email: guo-xiangyun@163.com.

reservoir water quality parameters[14]. These methods often fail to meet the requirements of multivariate prediction and nonlinear trends in water quality prediction tasks[15]. Dong et al.[16] proposed a room temperature forecasting model based on the Bezier curve equation.

Various machine learning methods, such as multiple linear regression, support vector machines, random forests, and more, are also used. Ma et al.[17] used a back propagation (BP) neural network model to predict water quality indicators at reservoir outflow stations. Otsuka et al.[18] proposed a seawater temperature prediction algorithm based on random forests, achieving a prediction accuracy of around 1°C. Usanapong and Boonnam, focusing on water quality prediction in seedling nurseries, tested models such as random forests, decision trees, logistic regression, and naive Bayes classification[19]. The results indicated that the random forest model performed the best in terms of training efficiency, testing efficiency, and accuracy. Li et al.[20] proposed a hybrid model based on multi-scale features for dissolved oxygen prediction in aquaculture. Nagaraju et al.[21] proposed soft computing coupled with wavelet analysis to predict ammonia content.

Compared with traditional time series models, machine learning delves deeper, enabling better fitting of complex patterns. However, traditional machine learning methods often rely on manual feature extraction. The relationships between different parameters and locations in aquaculture water quality are highly complex, posing a significant challenge for feature engineering[22-24]. In recent years, an increasing number of researchers have begun to use deep learning to investigate water quality prediction tasks. For instance, Li et al.[25] proposed a fusion prediction model, the Long Short-Term Memory (LSTM) combined with the Temporal Convolutional Network (TCN), known as the LSTM-TCN, to predict dissolved oxygen in aquaculture environments. LSTM is used to extract temporal features, handling long-term dependencies in complex time series, while the TCN establishes a fusion prediction model for time series data. Zhou et al.[26] employed deep neural network (DNN) and a proportional integral derivative residual elimination network (PID-RENet) to predict water quality. Li et al.[27] combined graph convolution with LSTM network, using time decomposition to predict river water quality, providing insights and methods for aquaculture water quality prediction.

However, the aforementioned prediction tasks suffer from the following shortcomings: 1) Focusing on predicting multiple variables at a single location or predicting a single variable at multiple locations. There is limited focus on tasks involving the prediction of multiple variables at multiple locations. The relationships of different water quality parameters in different locations are intricate. Extracting the mutual impact patterns of different indicators and locations is of great significance for predicting multiple variables at multiple locations. 2) Implicit extraction methods are often employed to understand the influences of different locations and variables, which weaken a model's ability to learn the relationships between different variables. 3) The interactive relationships between water quality variables often do not occur in real-time. Previous water quality predictions usually span a single time frame, making it challenging to extract temporal patterns across multiple time spans[28].

In recent years, spatiotemporal graph convolution models, due to their ability to simultaneously extract temporal and spatial features, have found wide applications in the spatiotemporal prediction domain. For example, Yu et al. and Feng et al.[29,30] employed a spatiotemporal graph convolutional network (STGCN) to predict traffic flow and speed. Li et al.[31] proposed a spatial-temporal graph convolutional network, which combined the graph convolution network with Informer to predict air quality. Xing et al.[32] employed it to predict the remaining useful life of the power transformer. Most existing studies utilize graph structures that heavily rely on predefined structures, wherein a fixed graph matrix is established based on prior knowledge of the spatial relationships between variables[33]. In water quality prediction for aquaculture, there are physically meaningful spatial relationships between water parameters at different depths and semantic spatial relationships between different parameters. These spatial relationships are intricate and challenging to capture, which makes the precise definition of a fixed graph matrix a formidable challenge.

Therefore, this study proposed an adaptive multi-channel spatiotemporal graph convolution model (AMTGCN). The model incorporates an adaptive graph construction layer, which is capable of integrating information gathered by different sensors and autonomously discerning the complex interdependencies among various water quality indicators, as well as between water quality indicators at different depths. Simultaneously, through multi-channel spatiotemporal graph convolution, the model extracts patterns of water quality across multiple periods. Additionally, the study explores the multi-step prediction performance of the proposed model, aiming to contribute insights into the development of the deep-sea aquaculture industry.

2 Materials and methods

2.1 Study area

The experimental marine ranch, Blue Diamond No. 2 large-scale ecological pen, is in Laizhou Bay (LZB), which is located in the south of the Bohai Sea and is the largest bay in Shandong Province of China (see Figure 1). LZB is a semi-enclosed shallow bay, with a mean depth of 9 m (maximum ~18 m), a coastline of 320 km, and a total area of about 700 000 hm^2[34]. Hence, the mariculture industry of Laizhou Bay is well-developed and an important aquaculture base in northern China. There are many mariculture farms along the eastern coast of Laizhou Bay, with various cultural modes and species[35]. However, the northeast-southeast currents dominate with a mean velocity of 20 cm/s, leading to a poor exchange of water[34], which negatively affects the aquaculture industry.

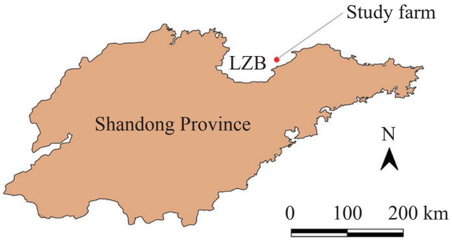


Figure 1 Overview of the location of the study farm in Shandong, China

The water quality data were collected from Blue Diamond No. 2 large-scale ecological pen, whose location is shown in Figure 1. It was installed and is operated by Laizhou Mingbo Aquatic Products Co., Ltd. Blue Diamond No. 2 facility has a diameter of 160 m and a depth of about 10 m, including a 20 000 m^3 nourishing water body. The breeding species is the spotted knifejaw (*Oplegnathus punctatus*), approximately 2 years old.

2.2 Data collection

Figure 2 shows the topological structure of the water quality data collection. Water was pumped up to the sensors and multi-parameter water-quality-monitoring instrument to obtain water quality values, which were then transmitted using the RS485 module and 4G network to a cloud platform and stored in a database[36]. Users could then read the water quality values via PC or mobile phone.

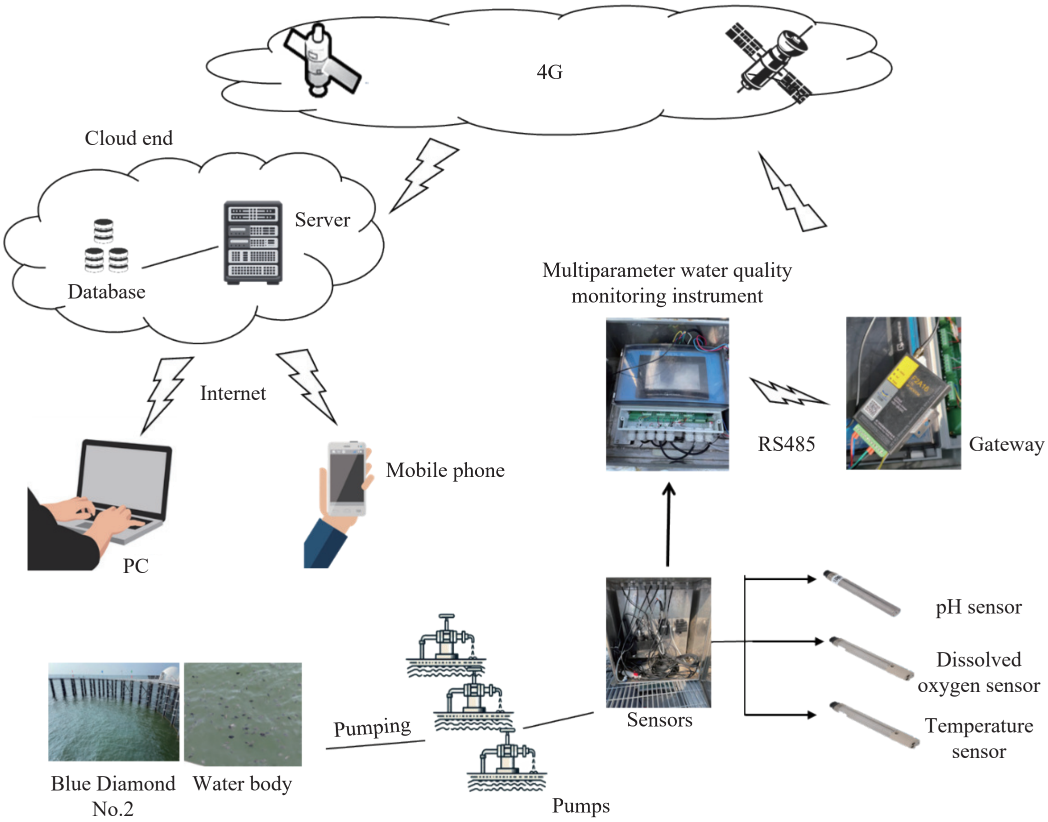


Figure 2 Topological structure of the data collection system

As pH, dissolved oxygen, and temperature are vital water quality parameters in aquaculture[37], they were collected in this research, and the sensors used here are listed in Table 1. All of them were produced by Shandong Dongrun Instrument Technology Co., Ltd, China. The water quality data were collected at three depths, 3 m, 6 m, and 9 m from the surface, as suggested by domain experts. At each of these depths, pH sensors, dissolved oxygen sensors, and temperature sensors were installed to monitor the water quality parameters. The water parameters at each depth were collected every 30 min. The collection period was from 19th August 2023 to 31st October 2023, this is the main breeding season for the spotted knifejaw in the marine ranch. Table 2 lists the statistical analysis of each parameter, from which it can be seen that the fluctuations in dissolved oxygen (DO) are greater than those in temperature, while the pH values remain almost unchanged across different depths. As depth increases, DO content gradually decreases, with a faster decline observed in the shallower layers compared to the deeper layers. The standard deviation of DO also decreases with depth, indicating smaller fluctuations in DO at greater depths. The data was collected during the summer, which explains why the temperature in the shallow layers is higher than in the deeper layers, with temperature variations also diminishing with increasing depth.

Table 1 Sensors used in this research

Sensors	Measurement range	Measurement accuracy	Measurement precision
DO	0-20 $\text{mg}\cdot\text{L}^{-1}$	±2%F.S	0.01 $\text{mg}\cdot\text{L}^{-1}$
Temperature	−10°C-55°C	±1%F.S	0.01°C
pH	0-14	±0.1	0.01

Table 2 Statistical analysis of each parameter

Metrics	DO/$\text{mg}\cdot\text{L}^{-1}$			Temperature/°C			pH		
	3 m	6 m	9 m	3 m	6 m	9 m	3 m	6 m	9 m
Min	5.92	5.91	5.85	15.60	15.60	15.10	8.09	8.07	7.93
Max	27.49	22.87	28.60	29.40	28.90	28.80	8.35	8.35	8.35
Average	9.49	9.26	9.25	23.53	23.51	23.51	8.23	8.23	8.22
Std.	1.33	1.14	1.26	3.69	3.65	3.62	0.05	0.05	0.05

2.3 Proposed model

2.3.1 Overall architecture of the model

The AMTGCN model comprises learning modules and an output module. The learning modules consist of multiple channels, each composed of a graph construction layer (GCL) and a spatiotemporal graph convolutional network layer (STGCL). The output module includes a fusion layer (FL), as illustrated in Figure 3. To effectively integrate information acquired from disparate sensors, this study has engineered a graph construction layer (GCL). This layer operates by treating the data collected from each sensor as a separate sequence, and it learns the interdependencies among these sequences to construct a graph structure matrix. This matrix reflects the spatial relationships between the water quality variables at various time spans and enhances the model's convergence speed. The STGCL employs multiple layers of recurrence to learn the spatiotemporal patterns of the water quality data, taking the graph structure matrix and water quality data as the input. The fusion layer integrates the information obtained from the different learning channels and predicts the water quality spatiotemporal data.

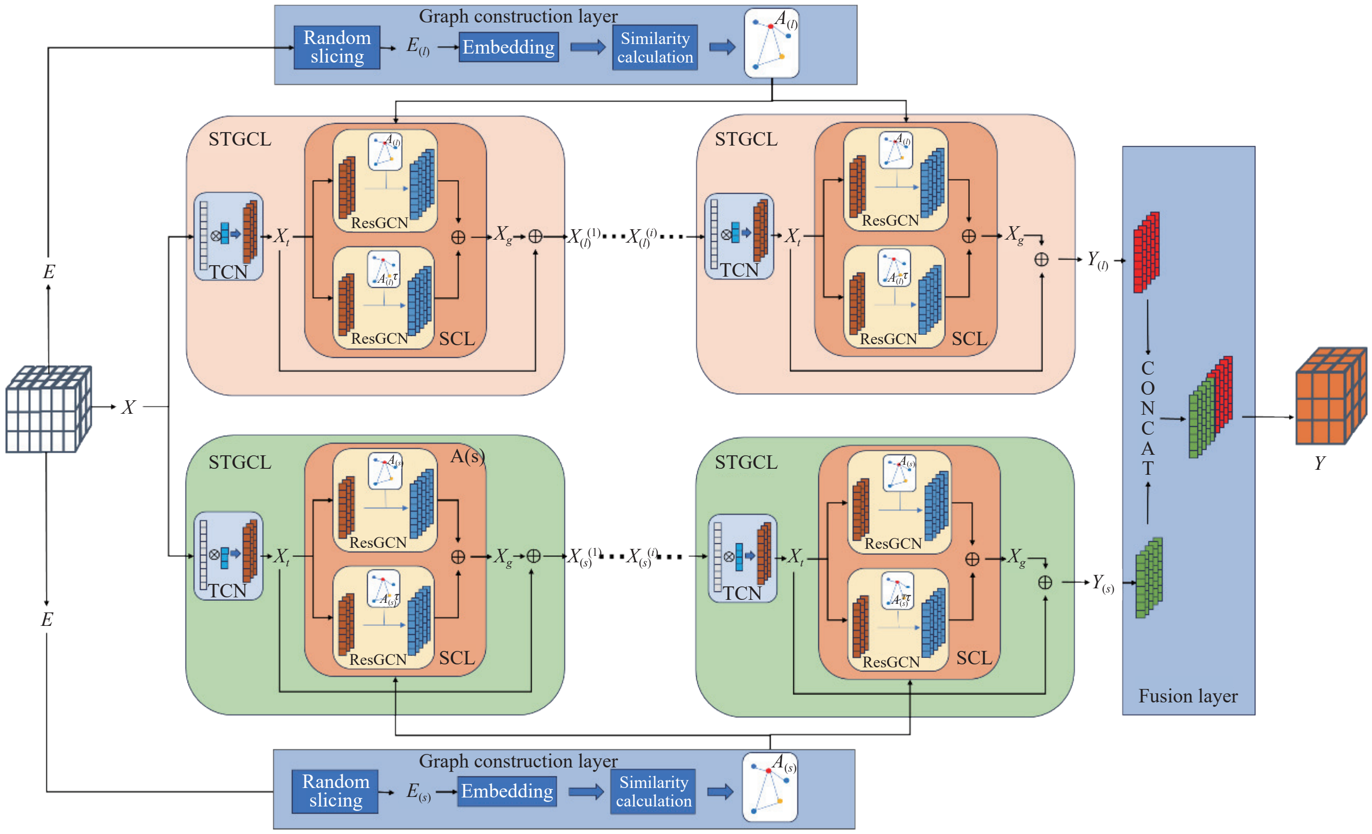


Figure 3 Overall structure of the model proposed in this study

In Figure 3, $\boldsymbol{E} \in \mathbb{R}^{n\times L}$ represents the entire dataset of the water quality sequence training data. $\boldsymbol{X} \in \mathbb{R}^{n\times l}$ corresponds to the water quality data sequence input into the model for each iteration, denoted as $\boldsymbol{X} \in \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \dots \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i \in \{x_i^1, x_i^2, x_i^3 \dots x_i^l\}, n \in \mathbb{R}$, $l \in \mathbb{R}$ and $i \in [1,n]$. Here, $n$is the number of water quality data sequences, totaling nine in this study, comprising three indicators (dissolved oxygen, temperature, pH) at three different depths. L represents the total length of the water quality training set sequences, and l is the length of the water quality data input into the model in each iteration, corresponding to the length of the input sequence, which is 12 in this research. $\boldsymbol{E}$ is the input into two graph construction layers, and the random slicing module divides it into random slices of different lengths in different channels, namely $\boldsymbol{E}_{(l)} \in \mathbb{R}^{n\times \text{long}}$ and $\boldsymbol{E}_{(s)} \in \mathbb{R}^{n\times \text{short}}$. $\boldsymbol{A}_{(l)}$, and $\boldsymbol{A}_{(s)} \in \mathbb{R}^{n\times n}$ are the graph structure matrices constructed by the graph construction layers for $\boldsymbol{E}_{(l)}$ and $\boldsymbol{E}_{(s)}$, respectively, representing the graph structure matrices at different time spans. $\boldsymbol{Y}_{(l)}$, and $\boldsymbol{Y}_{(s)} \in \mathbb{R}^{n\times\tau}$ denote the outputs of X, $\boldsymbol{A}_{(l)}$, and $\boldsymbol{A}_{(s)}$ after passing through the different channels of the STGCL. $Y \in \mathbb{R}^{n\times\tau}$ is the output of the predicted values, where τ is the model prediction step size.

2.3.2 Graph construction layer

In numerous previous studies, graph structures often heavily rely on predefined matrices, in which relationships between variables are predetermined based on prior knowledge, thus forming an invariant graph structure matrix. However, this study treated each water quality variable sequence as an individual node, in which it is impossible to precisely define the relationships between variables and predefine the graph structure matrix. Hence, this study proposed a graph construction layer (GCL), which computes the similarity between different variables over a specified time span based on the input data, thereby obtaining the graph structure matrix[38]. The graph construction layer is computed using Equations (1)-(3).

$$\boldsymbol{\alpha}_1 = sigmoid\left(\beta \boldsymbol{E}_{(i)} \boldsymbol{\theta}_1\right), \tag{1}$$

$$\boldsymbol{\alpha}_2 = sigmoid\left(\beta \boldsymbol{E}_{(i)} \boldsymbol{\theta}_2\right), \tag{2}$$

$$\boldsymbol{A} = topk\left(softMax\left(ReLU\left(\boldsymbol{\alpha}_1 \boldsymbol{\alpha}_2^{\mathrm{T}} / \sqrt[2]{d}\right)\right)\right), \tag{3}$$

where, $\boldsymbol{E}_{(i)}$ represents the randomly sliced water quality sequence data, where $\boldsymbol{E}_{(i)} \in \mathbb{R}^{n\times len}$, with n representing the number of water quality sequences and len denoting the length of the randomly sliced water quality sequence data. Parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are learnable, and β acts as a hyperparameter controlling the scaling of the sigmoid output. The outputs of the embedding layer are $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2 \in \mathbb{R}^{n\times d}$, where n represents the number of water quality sequence variables and d is the embedding dimension for the input water quality data. The $topk(\cdot)$ function retains the k largest values after calculating the dot product similarity, applying the ReLU and softMax activation functions, with all other values set to 0. The resulting matrix is the graph structure matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$.

This study establishes two graph construction layers (GCLs) for different time spans[39]. Both layers employ the same method and structure but differ in the length of the randomly sliced water quality sequence data. By controlling the length of the sliced data, the GCL can learn relationships across distinct time spans. Specifically, this study considers two time spans, a long span of 128 steps and a short span of 24 steps, resulting in corresponding long time span slices of $\boldsymbol{E}_{(l)} \in \mathbb{R}^{n\times \text{long}}$ and short time span slices of $\boldsymbol{E}_{(s)} \in \mathbb{R}^{n\times \text{short}}$, where $\text{long} = 128$ and $\text{short} = 24$.The two graph construction layers operate independently to learn the similarity of the information gathered by different sensors at 128- and 24-step intervals, respectively. Consequently, they produce graph structure matrices corresponding to these specific durations. These matrices are denoted as $\boldsymbol{A}_{(s)}$ and $\boldsymbol{A}_{(l)}$, which belong to the set of real-valued matrices of dimension $n\times n$, represented as $\boldsymbol{A}_{(s)}$ and $\boldsymbol{A}_{(l)} \in \mathbb{R}^{n\times n}$.

2.3.3 Spatiotemporal graph convolutional layer

The spatiotemporal graph convolutional layer (STGCL) serves as the model's core component, aiming to learn patterns in the water quality sequences across spatiotemporal dimensions. The STGCL comprises two components: the temporal convolutional network (TCN) for temporal convolution and the spatial convolutional layer (SCL) for spatial convolution, as shown in Figure 4. Through multiple recurrent layers, the STGCL comprehensively learns the spatiotemporal information in the water quality data. With the incorporation of multiple channels, it can capture spatiotemporal relationships at various scales. Distinct convolutional kernels and graph structures are possessed by different channels of the STGCL. Before entering the fusion layer, these layers operate independently, with data circulating multiple times within their specific spatiotemporal convolutional layers[40].

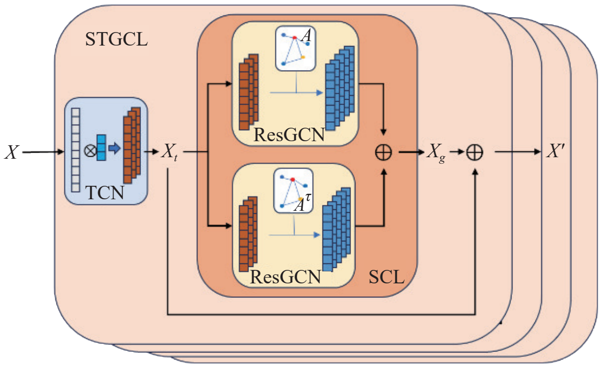


Figure 4 Spatiotemporal graph convolutional layer structure

Figure 4 shows the computation process of the STGCL. In this module, the input water quality sequence data $\boldsymbol{X}\in\mathbb{R}^{n\times l}$ undergoes a TCN. TCN utilizes one-dimensional convolutional kernels of varying sizes for different channels. The TCN output is then fed into spatial convolutional blocks to extract spatial features. Residual connections are employed between the TCN and the SCL to prevent model gradients from vanishing. Each STGCL iterates multiple times, extracting spatiotemporal features and passing them to the next layer of the STGCL. Stacking multiple layers enhances the model's learning of intricate spatiotemporal relationships, thereby improving prediction accuracy.

The SCL in Figure 4 consists of two residual graph convolutional networks (ResGCN), and each ResGCN utilizes a graph structure matrix $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ and its transposition $\boldsymbol{A}^{\mathrm{T}}\in\mathbb{R}^{n\times n}$ to more effectively capture one-directional influences among the variables[41]. In Figure 4, $\boldsymbol{X}_t\in\mathbb{R}^{n\times l'\times d^t}$ represents the output of the TCN, and $\boldsymbol{X}_g\in\mathbb{R}^{n\times l'\times d^g}$ represents the output of the SCL, where n is the number of variables, l' is the compressed length after the TCN, and d^t and d^g denote the output dimensions of the TCN and SCL, respectively.

The input to the STGCL consists of the input data $X\in\mathbb{R}^{n\times l}$ and the graph structure matrix $\boldsymbol{A}\in\mathbb{R}^{n\times n}$. After undergoing m iterations of spatiotemporal convolutional layers, the module outputs $\boldsymbol{X}'\in\mathbb{R}^{n\times(l-m+1)\times d^k}$, where n is the number of variables, l is the length of the input sequence, and d^k represents the output dimension size.

The temporal convolutional network (TCN) is utilized to handle time series data. Rooted in the concept of convolutional neural network (CNN), the TCN performs convolution operations along the time dimension and adeptly captures long-term dependencies within time series data, demonstrating a strong performance in dealing with extended sequences. By employing convolutional kernels of varying sizes, the TCN captures patterns at different temporal scales, enhancing its modeling capability for time series data. Structurally, the TCN is designed for efficient parallel computation, providing longer memory retention and higher parallelism than the traditional recurrent neural network (RNN). These attributes give the TCN advantages in processing long-term dependencies and large-scale data. The computation of the TCN is outlined in Figure 5 and Equation (4)[42].

$$\boldsymbol{X}_t=\boldsymbol{X}\otimes\boldsymbol{W}_g \tag{4}$$

where, $\boldsymbol{X}\in\mathbb{R}^{n\times l}$ represents the input into the TCN, where n is the number of input sequences and l is the length of the input sequence. The symbol ⊗ denotes the convolution operation, and $\boldsymbol{W}_g\in\mathbb{R}^{1\times\mathrm{kernel}}$ serves as the convolutional kernel for the TCN, with a size of $1\times\mathrm{kernel}$, constituting learnable parameters. $\boldsymbol{X}_t\in\mathbb{R}^{n\times(l-\mathrm{kernel}+1)}$ represents the output of the TCN. To ensure that the effective length of the information is consistent for each convolution in ResGCN, padding is not applied in the TCN section of this study. With each layer of the TCN, the data length undergoes compression, and the compression magnitude is (kernel−1).

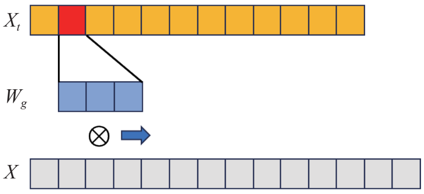


Figure 5 Diagram of the temporal convolutional network

Spatial convolutional layer: Graph convolution effectively utilizes information propagation relationships between nodes. In this study, various variable sequences are considered independent nodes. The application of graph convolution improves the comprehension and prediction of changes in water quality. Equation (5) represents the graph convolution (GCN)[43].

$$\boldsymbol{X}_g=g(\boldsymbol{X}_t)=\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{X}_t\boldsymbol{W} \tag{5}$$

where, $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ represents the graph structure matrix, $\boldsymbol{D}\in\mathbb{R}^{n\times n}$ is the degree matrix of $\boldsymbol{A}$, $\boldsymbol{I}\in\mathbb{R}^{n\times n}$ is the identity matrix, $\boldsymbol{X}_t\in\mathbb{R}^{n\times l}$ represents the input water quality sequence, and $\boldsymbol{W}$ stands for learnable parameters.

Traditional graph convolution experiences excessive smoothing, where gradients gradually converge to a fixed value, posing challenges for updating and impacting the model's learning effectiveness[44]. To address this, a novel graph convolution method was proposed: the residual graph convolutional network (ResGCN). The core concept is to incorporate the previous ResGCN operation results into the current one, creating a residual structure[45,46]. The ResGCN not only alleviates over-smoothing but also hastens model convergence and improves generalization. The computation equation for the ResGCN is provided in Equation (6).

$$\boldsymbol{X}_t^{(\mu)}=\boldsymbol{g}\left(\boldsymbol{X}_t^{(\mu-1)}\right)=\begin{cases}\left(r\boldsymbol{g}\left(\boldsymbol{X}_t^{(\mu-1)}\right)+(1-r)\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{g}\left(\boldsymbol{X}_t^{(\mu-1)}\right)\right)\boldsymbol{W}, & \mu>1\\ \left(r\boldsymbol{X}_t+(1-r)\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{X}_t\right)\boldsymbol{W}, & \mu=1\end{cases} \tag{6}$$

where, $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ represents the graph structure matrix; $\boldsymbol{D}\in\mathbb{R}^{n\times n}$ is the degree matrix of $\boldsymbol{A}$; $\boldsymbol{I}\in\mathbb{R}^{n\times n}$ is the identity matrix; $\boldsymbol{X}_t^{(i)}\in\mathbb{R}^{n\times l}$ represents the input water quality sequence for the i-th ResGCN iteration; when i=0, it denotes the output of the TCN; and $\boldsymbol{W}$ represents learnable parameters.

In Equation (6), $\boldsymbol{X}^{\mu-1} \in \mathbb{R}^{n\times l}$ denotes the output of the preceding ResGCN iteration. This method introduces a hyperparameter r, aiming to balance the weights between the previous and current ResGCN operations. When r=0, it degenerates to the conventional GCN formula, and when r=1, it transforms into a fully connected network. Adjusting the value of r controls the influence of the graph structure relationships on the model, adapting better to various application scenarios.

The ResGCN structure, as depicted in Figure 6, involves $\boldsymbol{X}_t \in \mathbb{R}^{n\times l}$ as the input for the ResGCN, while $\boldsymbol{X}_t^{(i)}$, and $\boldsymbol{H}_g^{(i)} \in \mathbb{R}^{n\times l}$ are the outputs of the ith ResGCN iteration. Here, $\boldsymbol{X}_t^{(i)}$ continues as the input for the next iteration, while $\boldsymbol{H}_g^{(i)}$ participates in the Concat operation. $\boldsymbol{X}_g \in \mathbb{R}^{n\times l}$ represents the output of the ResGCN. The Concat operation concatenates the outputs at each iteration, and the MLP serves to selectively fuse information using a 1×1 convolutional kernel, yielding the final ResGCN output.

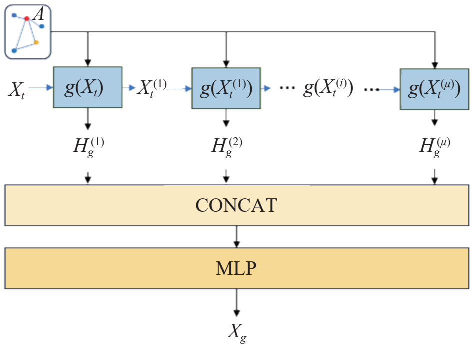


Figure 6 Residual graph convolution structure diagram

The model incorporates two channels for the GCL and STGCL, each sharing an identical structure. Both channels consist of their respective GCL and STGCL. By utilizing the GCL with different time spans, a temporal convolutional network with distinct kernel sizes, and varying numbers of iterations in the STGCL, the model aims to extract spatiotemporal relationships at different time scales.

2.3.4 Fusion layer

The fusion layer integrates information from various channels to predict future data by concatenating the outputs of different STGCLs with different scales. These concatenated outputs are then passed through a fully connected layer, generating the prediction result, as shown in Equation (7).

$$\boldsymbol{Y} = concat(\boldsymbol{Y}_{(l)}, \boldsymbol{Y}_{(s)})\boldsymbol{W} \tag{7}$$

where, $\boldsymbol{Y}_{(l)}$ and $\boldsymbol{Y}_{(s)} \in \mathbb{R}^{n\times\tau}$ represent the outputs from different channels, where n denotes the number of sequences in the water quality data, and τ represents the prediction step size. $\boldsymbol{W}$ represents learnable parameters, while concat(·) indicates the concatenation operation, and $\boldsymbol{Y} \in \mathbb{R}^{n\times\tau}$ is the final prediction result of the model.

2.3.5 Loss function

To comprehensively assess the impact of all data and improve prediction accuracy, the loss function calculation method involves initially computing the absolute errors for the predictions of all variable sequences separately. These absolute errors are then summed up, as illustrated in Equation (8)[47].

$$\text{loss} = \sum_{i=1}^{n}\left(\frac{1}{\tau}\sum_{j=1}^{\tau}\left|\hat{y}_j^i - y_j^i\right|\right) \tag{8}$$

where, n represents the number of sequences in the water quality data, τ is the prediction step size for each sequence, $\hat{y}_j^i$ denotes the predicted value of the model for the j^{th} time step of the ith sequence, while y_j^i represents the actual measurement value for the j^{th} time step of the i^{th} sequence.

2.3.6 Model parameters

Table 3 lists the parameter configurations for each layer in the model used in this study. Notably, the temporal convolutional network (TCN) and spatial convolutional layer (SCL) in the spatiotemporal graph convolutional layer alternate in operation.

Table 3 Model parameter configurations

Module	Number of iterations	Parameters
Graph construction layer (long)	1	Random slice length=128 Embedding dimensions=16 $\beta = 3$ d=16
Graph construction layer (short)	1	Random slice length=24 Embedding dimensions=16 $\beta = 3$ d=16
Spatiotemporal graph convolutional layer (long)	4	TCN kernel=1×4 r=0.5 TCN includes 4 hidden layers and the dimensions of each hidden layer are 16, 128, 64, and 1, respectively, SCL includes 3 hidden layers and the dimensions of each hidden layer are 64, 128, and 16, respectively.
Spatiotemporal graph convolutional layer (short)	4	TCN kernel = 1×2 r=0.5 TCN includes 4 hidden layers and the dimensions of each hidden layer are 16, 128, 64, and 1, respectively, SCL includes 3 hidden layers and the dimensions of each hidden layer are 64, 128, and 16, respectively.
Fusion layer	1	Hidden layer dimensions=16 Output layer dimensions=1

2.4 Evaluation metrics

To validate the accuracy of the model and compare it with LSTM, the TCN, the STGCN, the GCN-LSTM, and the single-channel AMTGCN (denoted as AMTGCN/w), this study employed the mean absolute error (MAE), as shown in Equation (9); root mean square error (RMSE), as shown in Equation (10); mean absolute percentage error (MAPE), as shown in Equation (11); and coefficient of determination (R^2), as shown in Equation (12), as evaluation metrics. To mitigate the impact of random errors, each model was independently trained three times, and the average of the evaluation metrics from the three runs was considered as the final calculation result.

$$\text{MAE} = \frac{1}{\tau\times m}\sum_{i=1}^{\tau\times m}|\hat{y}_i - y_i| \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{\tau\times m}\sum_{i=1}^{\tau\times m}(\hat{y}_i - y_i)^2} \tag{10}$$

$$\text{MAPE} = \frac{100\%}{\tau\times m}\sum_{i=1}^{\tau\times m}\frac{|\hat{y}_i - y_i|}{y_i} \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{\tau\times m}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{\tau\times m}(\bar{y} - y_i)^2} \tag{12}$$

where, τ represents the prediction time step, m denotes the number of prediction samples, $\hat{y}_i$ is the ith predicted value in the current sequence, y_i is the actual measurement value at the ith position in the current sequence, and $\bar{y}$ is the mean value of the actual measurements.

2.5 Comparison methods

To effectively evaluate the developed model and show the performance difference between different methods, current state-of-the-art methods were compared with the proposed model in this research. The methods in the comparative evaluation are as follows:

Long short-term memory (LSTM). A unique recurrent neural network (RNN) model with recurrent layers that is suitable for processing and predicting time series features with very long intervals and delays. Its network is effective for solving gradient disappearance and gradient explosion problems, which often occur in the traditional recurrent neural network[48].

Temporal convolutional network (TCN). A variant of the convolutional neural network (CNN) that combines the advantages of CNN and recurrent neural network, which is widely used for extracting spatial and temporal features of time series data for its dilated casual convolutions and expanded receptive field[49].

Spatial-temporal graph convolutional network (STGCN). It comprises several spatiotemporal convolutional blocks, which are a combination of graph convolutional layers[50] and convolutional sequence learning layers[51], to model spatial and temporal dependencies.

Graph convolutional network and long short-term memory network (GCN-LSTM). The combination of a graph convolutional network (GCN) and a long short-term memory (LSTM) network, in which the GCN extracts the spatial features of the non-Euclidean structure graph data, while the LSTM network extracts the temporal features of the data[52].

2.6 Experimental setup

The input of the model depends on the size of the sliding window. In this experiment, all data were segmented according to different sliding window sizes[51]. To ensure an adequate number of samples and sufficient information within each sample, the sliding window size ($\boldsymbol{w}$) was selected as $\boldsymbol{w} = \{13, 15, 18\}$. As shown in Figure 7, the blue segment represents the required input data length for the model ($l = 12$), and the red segment represents the model's output. In this study, multi-step prediction was performed with output step sizes $\boldsymbol{\tau} = \{1, 3, 6\}$.

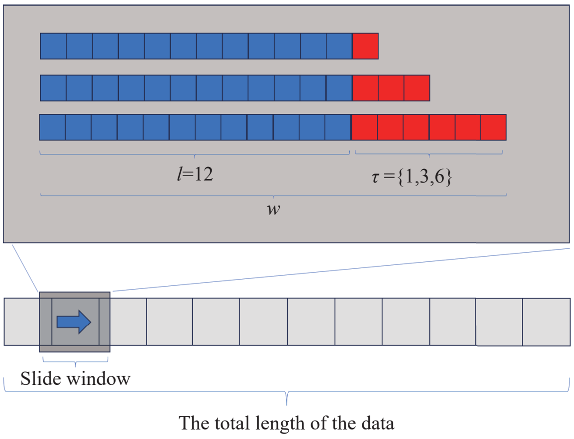


Figure 7 Sliding window demonstration

In this study, the dataset was partitioned into training, validation, and test sets in a ratio of 7:1:2. The programming language used was Python 3.9, and the deep learning framework employed was PyTorch. The training process of the model utilized the Adam optimizer with an initial learning rate of 0.01. The learning rate was dynamic, decreasing by a factor of 0.1 every 20 iterations. The batch size was set to 256, and the model iteration count was fixed at 100 during training.

3 Results and discussion

3.1 Loss line analysis

The loss function is used to express the degree of difference between the forecast and actual data. The smaller the loss function, the better the robustness of the model[6]. Figure 8 shows the loss function graph for our AMTGCN and the comparison models. It can be seen that training loss decreased for all models and converged after about 40 iterations. The TCN converged earlier than the AMTGCN, but its loss fluctuated more than the AMTGCN, which shows that it may be overfitting the test data. The loss of the LSTM, STGCN, and GCN-LSTM models were all greater than the AMTGCN, and the loss of the STGCN significantly fluctuated, which shows that the network is very unstable for water quality prediction in this situation.

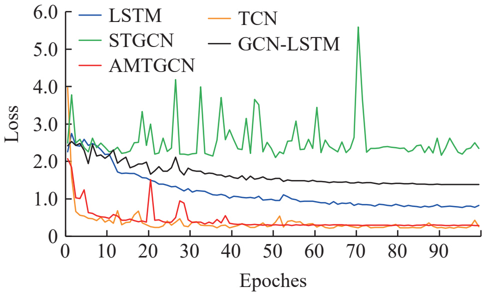
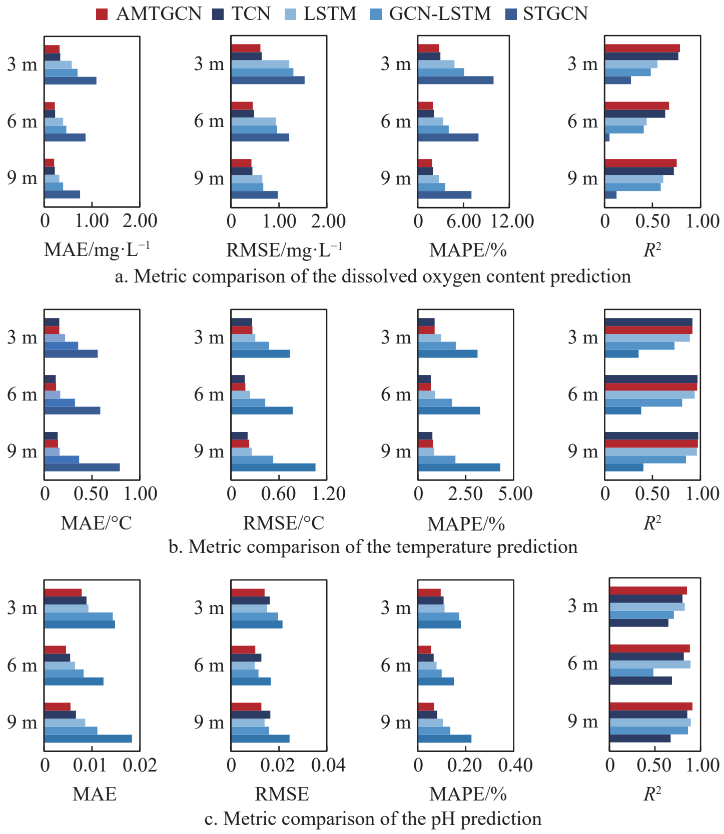


Figure 8 Loss function of all the models

3.2 Single-step prediction analysis

The MAE, RMSE, MAPE, and R^2 metrics were employed to analyze the model's performance. It is worth mentioning that the smaller the MAE, RMSE, and MAPE values are, the better the performance of the model, and vice versa. On the contrary, a larger R^2 value suggests a better prediction performance.

All the prediction metrics are shown in Figure 9. The red bar represents the AMTGCN, and the blue bars represent the comparison models. The AMTGCN has the smallest MAE, RMSE, and MAPE values and the biggest R^2 in the dissolved oxygen content and pH prediction. However, the AMTGCN is slightly worse than the TCN in temperature prediction, with MAE, RMSE, MAPE, and R^2 values being 0.1599°C, 0.2675°C, 0.8783%, and 0.9154, respectively, while the four metrics of the TCN are 0.1596°C, 0.2635°C, 0.8767%, and 0.9179, respectively. Combined with Figures 10-12, which show a comparison between the prediction values and the ground truth (to clearly see the difference between the prediction value and the observation value, some parts were enlarged), the trends in the DO and pH vary much more than the temperature trend. This result may be because the AMTGCN, which integrates the TCN and ResGCN, can capture more fluctuation information. From a depth perspective, as the depth increases, the prediction performance for dissolved oxygen improves. However, there is no apparent pattern observed in the prediction results for temperature and pH. Table 4 lists the mean of the metrics at three depths. In this experiment, the MAE, RMSE, MAPE, and R^2 of the AMTGCN prediction show an average improvement of 35.34%, 39.41%, 34.59%, and 10.01% compared to LSTM, respectively; an average improvement of 74.39%, 65.29%, 74.34%, and 113.71% compared to the STGCN, respectively; an average improvement of 54.48%, 50.02%, 53.26%, and 30.20% compared to GCN-LSTM, respectively; and an average improvement of 3.31%, 2.26%, 4.16%, and 3.33% compared to the TCN, respectively.



a. Metric comparison of the dissolved oxygen content prediction

b. Metric comparison of the temperature prediction

c. Metric comparison of the pH prediction

Figure 9 Comparison of the models in predicting multi-parameters and multi-positions

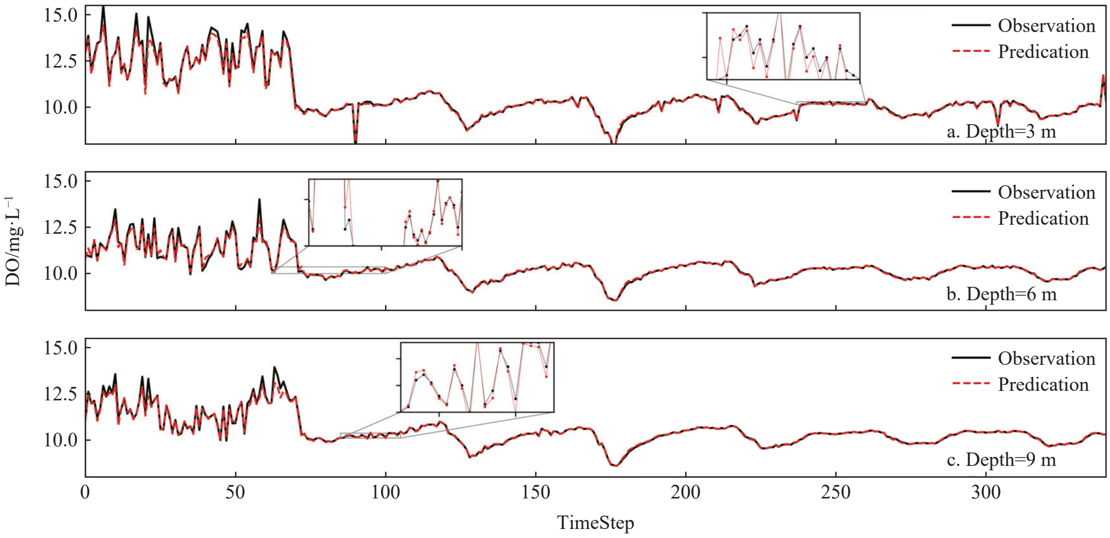


Figure 10 Comparison of the AMTGCN prediction and observation values of the DO

The LSTM and GCN-LSTM models are based on recurrent neural network (RNN), which has advantages in handling sequential data. However, from the experimental results, it is evident that the TCN-type models (AMTGCN, TCN) perform significantly better than the RNN-type models in prediction tasks[47]. This may be attributed to the stronger ability of the TCN-type models to extract spatial relationships between multivariate time series data. In contrast, the RNN-type models are relatively weaker in learning such relationships, leading to a suboptimal overall performance[53].

The comparison between the TCN, STGCN, and AMTGCN also reveals differences between the models without graphs, models with graphs, and models with adaptive graphs. For data such as water quality data without clear physical graph structures, an incorrect graph structure may actually degrade the prediction performance. An accurate graph structure is helpful for learning the spatial relationships between data. Therefore, the adaptive graph structure of the AMTGCN performs better in prediction than the TCN or STGCN. Additionally, the STGCN referenced in this study was initially used for traffic prediction, where there are typically clear graph structures. This might be a reason for the less-than-ideal prediction results.

Further comparison between the TCN and AMTGCN experimental results indicates that establishing an adaptive graph can effectively enhance the model’s prediction accuracy when

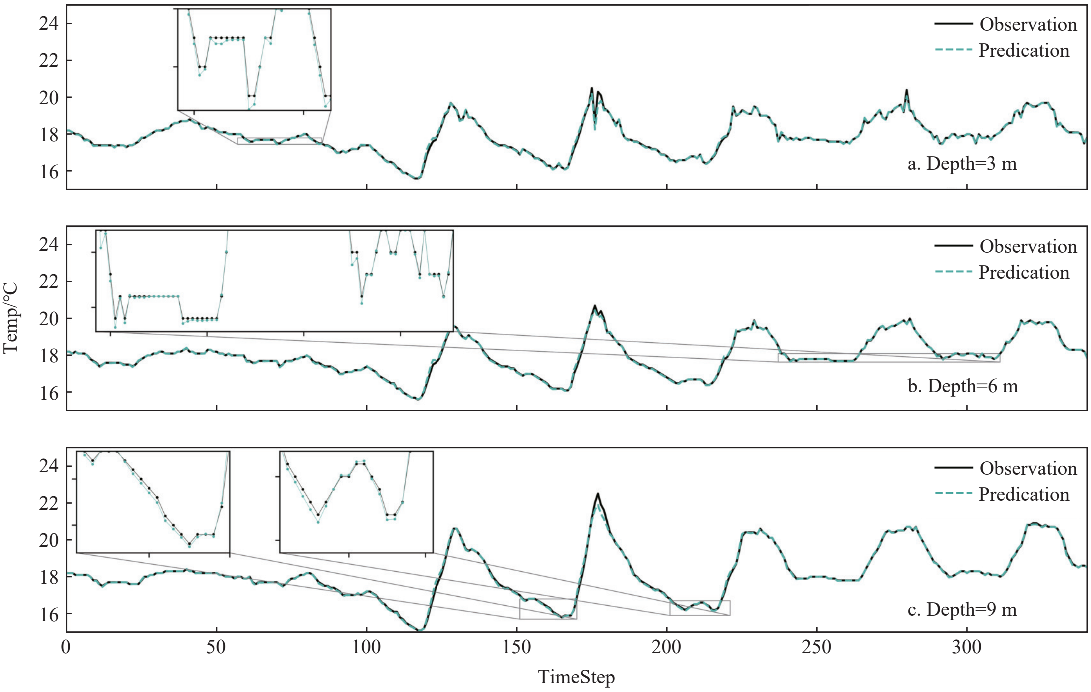


Figure 11 Comparison of the AMTGCN prediction and observation values of the water temperature

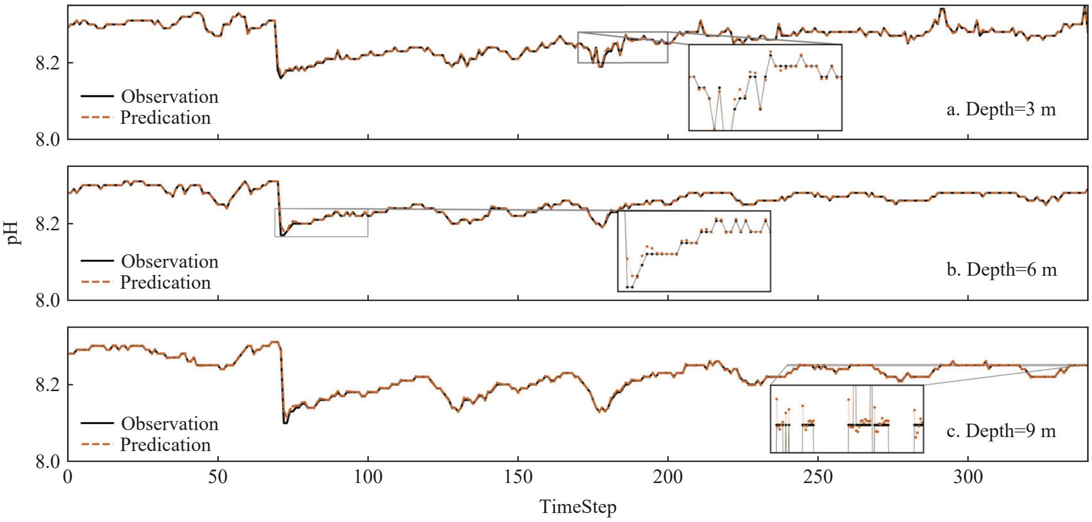


Figure 12 Comparison of the AMTGCN prediction and observation values of the water pH

Table 4 Error analysis between the AMTGCN and the other comparison methods

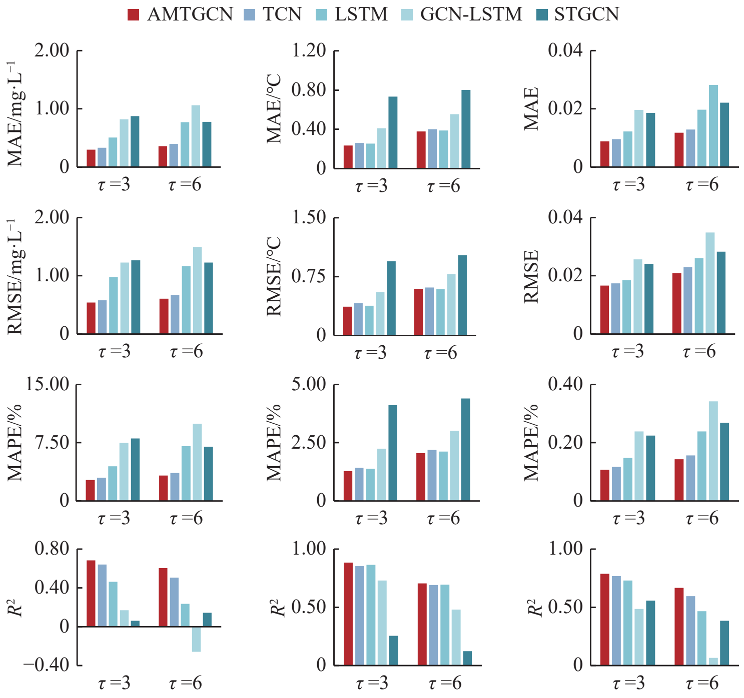
Variables	Metrics	AMTGCN	TCN	LSTM	GCN-LSTM	STGCN
DO	MAE/mg·L^{-1}	**0.2513**	0.2661	0.4291	0.5206	0.9030
	RMSE/mg·L^{-1}	**0.4995**	0.5242	0.9349	0.9804	1.2455
	MAPE/%	**2.2394**	2.3728	3.6407	4.5797	8.3164
	R^2	**0.7359**	0.7075	0.5333	0.4914	0.1520
Temperature	MAE/°C	0.1434	**0.1413**	0.1826	0.3485	0.6463
	RMSE/°C	0.2252	**0.2148**	0.2685	0.4787	0.857
	MAPE/%	0.7828	**0.7718**	0.9926	1.9059	3.5607
	R^2	0.9514	**0.9547**	0.9316	0.7970	0.3807
pH	MAE	**0.0060**	0.0070	0.0081	0.0112	0.0152
	RMSE	**0.0123**	0.0151	0.0130	0.0156	0.0209
	MAPE/%	**0.0730**	0.0848	0.0983	0.1365	0.1845
	R^2	**0.8815**	0.8237	0.8701	0.6844	0.6693

dealing with data with different spatial distributions and mutual influences. This suggests that, compared to the TCN, the model of this study exhibits better adaptability and predictive capabilities for data with different feature distributions and interaction patterns.

3.3 Multi-step prediction analysis

Most traditional environmental prediction methods are designed for single-step prediction, which plays a limited role in aquaculture farm management and early warning[47]. Hence, multi-step prediction is more meaningful than single-step prediction, and investigation of a model's accuracy and time step relationship can also evaluate a model's capacity for long time periods[54].

In this research, the collected data was used to assess the accuracy of the AMTGCN at three distinct time steps: single-step prediction, three-step prediction, and six-step prediction. The duration of each was 30 min, 1.5 h, and 3 h, respectively. As the one-step prediction results are shown in Section 3.2, Figure 13 only illustrates the comparison of the multi-step (three-step and six-step) prediction errors of the three parameters. This study employed the averaged MAE, RMSE, MAPE, and R^2 at the three depths to assess the performances of each model in predicting the DO concentration, temperature, and pH. The metric results are shown in their respective columns. Additionally, the red bar represents the result of the AMTGCN. Figure 13 shows that the AMTGCN outperformed the TCN, LSTM, GCN-LSTM, and STGCN in all three of the water quality parameters.



a. DO b. Temperature c. pH

Figure 13 Comparison of the multi-step prediction accuracy of the three parameters

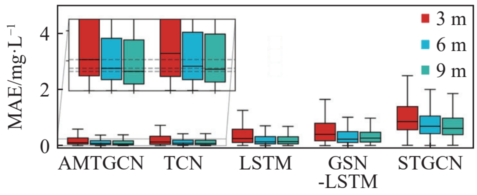
Specifically, in the three-step predictions, the MAE, RMSE, MAPE, and R^2 of the AMTGCN results showed an average improvement of 30.06%, 32.73%, 31.79%, and 14.40% compared to LSTM, respectively; an average improvement of 66.73%, 58.58%, 67.13%, and 169.74% compared to the STGCN, respectively; an average improvement of 56.68%, 48.69%, 59.04%, and 69.81% compared to the GCN-LSTM, respectively; and an average improvement of 9.94%, 8.24%, 10.05%, and 4.00% compared to the TCN, respectively. In six-step predictions, the MAE, RMSE, MAPE, and R^2 of the AMTGCN results showed an average improvement of 36.62%, 31.64%, 41.78%, and 28.70% compared to LSTM, respectively; an average improvement of 53.40%, 46.47%, 52.99%, and 176.02% compared to STGCN, respectively; an average improvement of 54.60%, 47.31%, 58.73%, and 526.99% compared to GCN-LSTM, respectively; and an average improvement of 7.91%, 6.46%, 8.07%, and 0.07% compared to the TCN, respectively. Considering three experiments (single-, three-, and six-step), the AMTGCN results showed an average improvement of 34.01%, 34.59%, 36.05%, and 17.71% compared to LSTM, respectively; an average improvement of 64.84%, 56.78%, 64.82%, and 153.16% compared to the STGCN, respectively; an average improvement of 55.25%, 48.67%, 57.01%, and 209.00% compared to GCN-LSTM, respectively; and an average improvement of 7.05%, 5.66%, 7.42%, and 2.47% compared to the TCN, respectively.

It is worth noting that the GCN-LSTM model prediction performance is poor, with a minus R^2 value when τ=6 for DO prediction. This is closely related to the nonlinearity and uncertainty of the DO content itself, and the GCN-LSTM could not effectively capture the characteristics of the variation, leading to a low prediction accuracy.

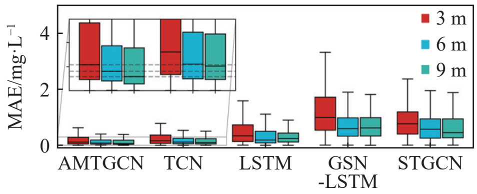
The variance of the DO content in the first 3010 data points was 0.77, and the mean value was 9.06 mg/L, while the variance in the last 522 data points was 3.15, and the mean value was 10.88 mg/L; the distribution characteristics of the DO time series data shows a large change. Compared with the other two parameters, the DO content displays a different change pattern over time due to its own characteristics. The DO changed more steadily in the former period and greatly fluctuated in the later period, which was a large change from the previous series change pattern. Changes in the water quality indicators will be affected by the environment, climate, and other factors, which will undoubtedly increase the forecasting difficulty. The time series processing methods, such as Empirical Mode Decomposition (EMD) and wavelet transformation, may be used to conduct time-series predictions based on different scales and frequency features to extract features, thus improving the prediction performance[55].

3.4 Error analysis

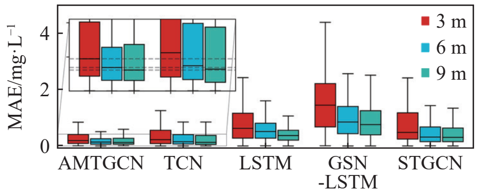
The comparative error distribution of each model is illustrated in Figures 14, 15, and 16. From a depth perspective, it was observed that as the depth increases, the prediction MAE range of all the models for dissolved oxygen becomes narrower, as shown in Figure 14. The temperature prediction error distribution remained relatively invariant, with the exception of the STGCN model, which exhibited a broader error range in one-step predictions, as illustrated in Figure 15. pH prediction errors showed relatively small ranges at a depth of 6 m compared with depths of 3 m and 9 m, as evidenced in Figure 16. With increasing prediction steps, there is a concomitant degradation in the performance of all the models. In these experiments, the AMTGCN model provides a small MAE and narrow MAE range, standing out for its robustness.



a. MAE metric distribution in single-step prediction (30 min)

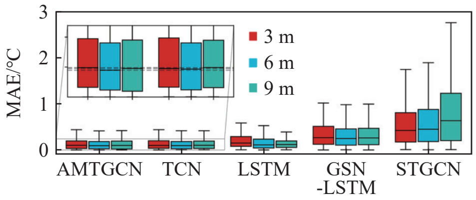


b. MAE metric distribution in three-step prediction (1.5 h)

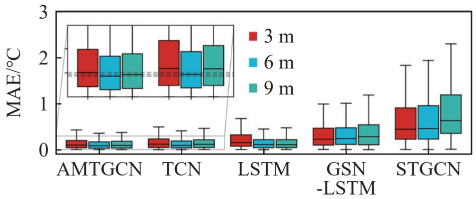


c. MAE metric distribution in six-step prediction (3 h)

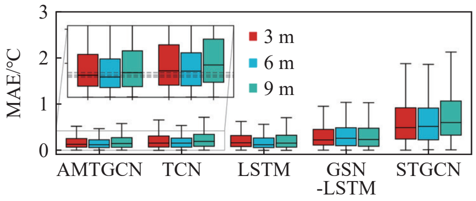
Figure 14 MAE analysis of the DO prediction at each step



a. MAE metric distribution in single-step prediction (30 min)



b. MAE metric distribution in three-step prediction (1.5 h)

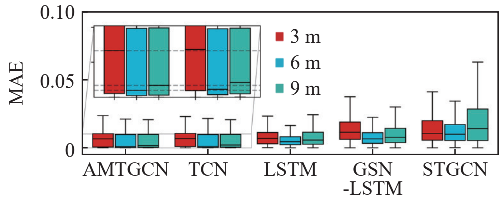


c. MAE metric distribution in six-step prediction (3 h)

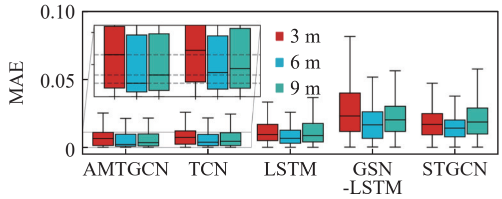
Figure 15 MAE analysis of temperature prediction at each step

The error distribution can also intuitively reflect the stability and accuracy of the model's predictions. Comparing the three models that utilized graph convolution - AMTGCN, GCN-LSTM, and STGCN - it can be observed that the maximum errors of the GCN-LSTM and STGCN are consistently the highest, indicating that these two models performed poorly in predicting the experimental data.

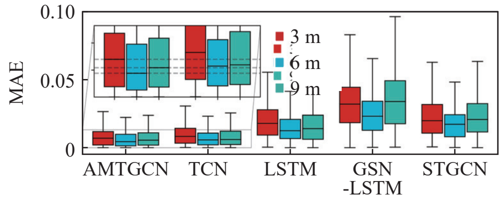
From the perspective of stability, the error distribution of the GCN-LSTM and STGCN models is not concentrated and generally larger, while the error distribution of the AMTGCN model is concentrated and centered at lower levels, indicating better model stability compared to the other models. Regarding generalization, although the errors of all models inevitably increase with increasing steps, the AMTGCN remained in the optimal position, demonstrating superior performance. The relatively small variation in the error distribution of the AMTGCN indicates better generalization, whereas the larger fluctuations in the GCN-LSTM and STGCN models suggest poorer generalization.



a. MAE metric distribution in single-step prediction (30 min)



b. MAE metric distribution in three-step prediction (1.5 h)



c. MAE metric distribution in six-step prediction (3 h)

Figure 16 MAE analysis of pH prediction at each step

The performances of the AMTGCN, TCN, and LSTM models are quite comparable. Even when predicting pH at a depth of 6 m in a single step, the maximum absolute error of the LSTM model is smaller than that of the AMTGCN. However, as the prediction step increases, LSTM exhibits greater fluctuations, indicating that it is less generalized compared to the AMTGCN. The error distribution of the AMTGCN and TCN is very close, but the AMTGCN outperforms the TCN in terms of the maximum error, error concentration, and median, suggesting that the AMTGCN is more effective than the TCN, as shown in the enlarged box.

3.5 Ablation study

Table 5 lists the ablation experimental results on the multi-channel STGCL to validate its essential properties. Here, the AMTGCN/w represents the single-channel STGCL model. The metric values are the average at the three depths. As indicated in Table 5, in single-step prediction, the AMTGCN outperforms the AMTGCN/w when predicting pH, which may be because pH changes subtly compared to DO and temperature. In multi-step prediction (τ=3, 1.5 h; τ=6, 3 h), the AMTGCN had a better prediction performance than the AMTGCN/w in all three parameters. Comparing the AMTGCN/w with the AMTGCN, when the prediction step was a single step, the AMTGCN/w demonstrated local superiority over the AMTGCN. However, when the prediction step exceeded a single step, the performance of the AMTGCN surpassed that of the AMTGCN/w. This may suggest that with an increase in the prediction scale, different time channels spanning various durations can effectively enhance the model's predictive capabilities.

Table 5 Comparison of the AMTGCN ablation experiment in water quality prediction

Step	Metrics	DO		Temp.		pH	
		AMTGCN/w	AMTGCN	AMTGCN/w	AMTGCN	AMTGCN/w	AMTGCN
τ=1	MAE	**0.2495**	0.2513	**0.1383**	0.1434	0.0062	**0.0060**
	RMSE	**0.4980**	0.4995	**0.2165**	0.2252	0.0124	**0.0123**
	MAPE/%	**2.2220**	2.2394	**0.7537**	0.7828	0.0751	**0.0730**
	R^2	**0.7374**	0.7359	**0.9541**	0.9514	0.8807	**0.8815**
τ=3	MAE	0.3036	**0.2980**	0.2421	**0.2345**	0.0092	**0.0088**
	RMSE	0.5583	**0.5434**	0.3724	**0.3660**	0.0173	**0.0166**
	MAPE/%	3.3094	**2.6952**	2.1260	**1.2785**	0.1446	**0.1072**
	R^2	0.6665	**0.6824**	0.8811	**0.8851**	0.7700	**0.7875**
τ=6	MAE	0.3600	**0.3572**	0.3903	**0.3784**	0.0119	**0.0118**
	RMSE	0.6054	**0.6052**	0.6042	**0.5945**	0.0209	**0.0209**
	MAPE/%	3.3095	**3.2833**	2.1260	**2.0595**	0.1446	**0.1434**
	R^2	0.6056	**0.6058**	0.6965	**0.7062**	0.6652	**0.6659**

4 Limitations

There are some limitations in this research. As the sensors are sensitive to the ocean environment and easily corroded by marine water, the water quality parameter was tested by pumping marine water up to the multi-parameter water-quality-monitoring instrument through a tube, which led to a low sampling frequency and an insufficient amount of available data to facilitate the examination of more sophisticated maps.

Because the culture season in the offshore large-scale aquaculture pen is from July to October, and some time is needed to debug the monitoring and data collection system, the data used does not cover the whole culture season. At the same time, because of the cost of the experiment, weather data was not collected nor incorporated into the developed model.

5 Conclusions and future research

Water quality is of great significance for aquaculture, which is characterized by complex nonlinearity and instability, and water quality variables often exhibit complex correlations with each other. Hence, it is difficult to precisely predict them and further support aquaculture management. This study proposed a predictive model based on spatiotemporal convolution, introducing adaptive graphs and multi-channel spatiotemporal convolution layers to enhance prediction accuracy. The adaptive graph can automatically extract relationships between variables to construct a graph matrix, eliminating the need for a predefined graph matrix. Comparative experiments with LSTM, TCN, STGCN, and GCN-LSTM models revealed that the AMTGCN exhibits superior performance in predicting water quality tasks with complex water quality data. This suggests that this model yields better results when dealing with data possessing intricate structures and unclear structural relationships, offering a new approach and perspective for spatiotemporal predictions of water quality data. Finally, the model can be extended to other similar sequential prediction domains.

In the future, efforts will be dedicated to optimizing the AMTGCN for higher prediction accuracy and lower model complexity. This may involve exploring more efficient time convolution methods and precise adaptive graph construction algorithms to expedite model training and enhance its effectiveness. Additionally, considering the relatively small dataset in this study, validation and optimization on larger datasets can be explored to improve the model's generalization performance. Finally, weather data will be collected together with water quality parameters to further improve the prediction performance.

Acknowledgements

This work was funded by the National Key Research and Development Program of China: Sino-Malta Fund 2022 "Autonomous Biomimetic Underwater Vehicle for Digital Cage Monitoring" (Grant No. 2022YFE0107100).

[References]

[1] Price C, Black K D, Hargrave B T, Morris Jr J A. Marine cage culture and the environment: Effects on water quality and primary production. Aquaculture Environment Interactions, 2015; 6: 151–174.

[2] Garlock T, Asche F, Anderson J, Ceballos-Concha A, Love D C, Osmundsen T C, et al. Aquaculture: The missing contributor in the food security agenda. Global Food Security, 2022; 32: 100620.

[3] Nowakowski A J, Canty S W J, Bennett N J, Cox C E, Valdivia A, Deichmann J L, et al. Co-benefits of marine protected areas for nature and people. Nature Sustainability, 2023; 6: 1210–1218.

[4] Wang F L, Gao Y T, Gao Y H, Li M Y, Zhang B Z, Guan C T, et al. Preliminary study on land-sea alternate aquaculture of spotted knifejaw *Oplegnathus punctatus*. Fishery Modernization, 2022; 49(4): 8–14. (in Chinese)

[5] Lu H Y, Cheng C Y, Cheng S C, Cheng Y H, Lo W C, Jiang W L, et al. A low-cost AI buoy system for monitoring water quality at offshore aquaculture cages. Sensors, 2022; 22(11): 4078.

[6] Huan J, Li H, Li M B, Chen B. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Chang Zhou fishery demonstration base, China. Computers and Electronics in Agriculture, 2020; 175: 105530.

[7] Hu Z H, Zhang Y R, Zhao Y C, Xie M S, Zhong J Z, Tu Z G, et al. A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. Sensors, 2019; 19(6): 1420.

[8] Meyer A M, Klein C, Fünfrocken E, Kautenburger R, Beck H P. Real-time monitoring of water quality to identify pollution pathways in small and middle scale rivers. Science of The Total Environment, 2019; 651(Part2): 2323–2333.

[9] Li D S, Zhao W J, Hu J Z, Zhao S W, Liu S E. A long-term water quality prediction model for marine ranch based on time-graph convolutional neural network. Ecological Indicators, 2023; 154: 110782.

[10] Lin W, Luo H M, Wu J Y, Hung T-C, Cao B B, Liu X L, et al. A review of the emerging risks of acute ammonia nitrogen toxicity to aquatic decapod crustaceans. Water, 2023; 15(1): 27.

[11] Erickson R J. An evaluation of mathematical models for the effects of pH and temperature on ammonia toxicity to aquatic organisms. Water Research, 1985; 19(8): 1047–1058.

[12] Eze E, Halse S, Ajmal T. Developing a novel water quality prediction model for a South African aquaculture farm. Water, 2021; 13(13): 1782.

[13] Zheng Z Q, Ding H, Weng Z, Wang L X. Research on a multiparameter water quality prediction method based on a hybrid model. Ecological Informatics, 2023; 76: 102125.

[14] Wang J, Zhang L Y, Zhang W, Wang X D. Reliable model of reservoir water quality prediction based on improved ARIMA method. Environmental Engineering Science, 2019; 36(9): 1041–1048.

[15] Phan T T H, Nguyen X H. Combining statistical machine learning models with ARIMA for water level forecasting: The case of the red river. Advances in Water Resources, 2020; 142: 103656.

[16] Dong Q X, Liu J C, Qu M. Simple model for predicting hourly air temperatures inside Chinese solar greenhouses. Int J Agric & Biol Eng, 2023; 16(5): 56–60.

[17] Ma L L, Zhou L F, Wang T L. Water quality prediction based on BP neural network at Dahuofang Reservoir, China. Nature Environment and Pollution Technology, 2015; 14(3): 727–732.

[18] Otsuka T, Kitazawa Y, Ito T. Seawater temperature prediction method for sustainable marine aquaculture. Preprints. Available: https://doi.org/10.20944/preprints201709.0114.v1. Accessed on [2023-11-10].

[19] Usanapong R, Boonnam N. Data standardization analysis for water quality parameters of nursery aquaculture. In: 2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Prachuap Khiri Khan: IEEE, 2022; pp.1–4. doi: 10.1109/ECTI-CON54298.2022.9795389.

[20] Li C, Li Z, Wu J, Zhu L, Yue J. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. Information Processing in Agriculture, 2018; 5(1): 11–20.

[21] Nagaraju T V, Sunil B M, Chaudhary B, Prasad C D, Gobinath R. Prediction of ammonia contaminants in the aquaculture ponds using soft computing coupled with wavelet analysis. Environmental Pollution, 2023; 331: 121924.

[22] Ahmed U, Mumtaz R, Anwar H, Shah A A, Irfan, R, García-Nieto J. Efficient water quality prediction using supervised machine learning. Water, 2019; 11(11): 2210.

[23] Zhu N Y, Liu X, Liu Z Q, Hu K, Wang Y K, Tan J L, et al. Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. Int J Agric & Biol Eng, 2018; 11(4): 32–44.

[24] Duan E Z, Wang L J, Wang H Y, Hao H Y, Li R L. Remaining feed weight estimation model for health monitoring of meat rabbits based on deep convolutional neural network. Int J Agric & Biol Eng, 2022; 15(1): 233–240.

[25] Li W S, Wei Y G, An D, Jiao Y S, Wei Q. LSTM-TCN: Dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. Environmental Science and Pollution Research, 2022; 29: 39545–39556.

[26] Zhou X H, Wang J P, Liu Y R, Duan Q L. Deep learning with PID residual elimination network for time-series prediction of water quality in aquaculture industry. Computers and Electronics in Agriculture, 2023; 212: 108125.

[27] Li Y L, Zhang L, Li L. Study on water quality prediction of south source of Qiantang River based on GCN-LSTM. Yellow River, 2023; 45(12): 83–87, 95. (in Chinese)

[28] Skarding J, Gabrys B, Musial K. Foundations and modelling of dynamic networks using dynamic graph neural networks: A survey. IEEE Access, 2021; 9: 79143–79168.

[29] Yu B, Yin H T, Zhu Z X. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018; pp.3634–3640. doi: 10.24963/ijcai.2018/505.

[30] Feng A, Tassiulas L. Adaptive graph spatial-temporal transformer network for traffic forecasting. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, ACM: Atlanta: ACM, 2022; pp.3933–3937.

[31] Li P F, Zhang T, Jin Y T. A spatio-temporal graph convolutional network for air quality prediction. Sustainability, 2023; 15(9): 7624.

[32] Xing M D, Ding W L, Zhang T P, Li H. STCGCN: A spatio-temporal complete graph convolutional network for remaining useful life prediction of power transformer. International Journal of Web Information Systems, 2023; 19(2): 102–117.

[33] Dwivedi V P, Luu A T, Laurent T, Bengio Y, Bresson X. Graph neural networks with learnable structural and positional representations. arXiv. Available: https://arxiv.org/abs/2110.07875v2. Accessed on [2023-11-10].

[34] Huang Q, Olenin S, Sun S, De Troch M. Impact of farming non-indigenous scallop *Argopecten irradians* on benthic ecosystem functioning: A case-study in Laizhou Bay, China. Aquaculture Environment Interactions, 2018; 10: 227–241.

[35] Han Q F, Zhang X R, Xu X Y, Wang X L, Yuan X Z, Ding Z J, et al. Antibiotics in marine aquaculture farms surrounding Laizhou Bay, Bohai Sea: Distribution characteristics considering various culture modes and organism species. Science of The Total Environment, 2021; 760: 143863.

[36] Luo H P, Li G L, Peng W F, Song J, Bai Q W. Real-time remote monitoring system for aquaculture water quality. Int J Agric & Boil Eng, 2015; 8(6): 136–143.

[37] Rasheed Abdul Haq K P, Harigovindan V P. Water quality prediction for smart aquaculture using hybrid deep learning models. IEEE Access, 2022; 10: 60078–60098.

[38] Wu Z H, Pan S R, Long G D, Jiang J, Chang X J, Zhang C Q. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, 2020; pp.753–763. doi: 10.1145/3394486.3403118.

[39] Chen Y Z, Xie Z X. Multi-channel fusion graph neural network for multivariate time series forecasting. Journal of Computational Science, 2022; 64: 101862.

[40] Fang Y C, Ren K, Shan C H, Shen Y F, Li Y, Zhang W N, et al. Learning decomposed spatial relations for multi-variate time-series modeling. In: AAAI-23 Technical Tracks 6 - AAAI Technical Track on Machine Learning I, 2023; 37(6): 7530–7538.

[41] Klicpera J, Bojchevski A, Günnemann S. Predict then propagate: Graph neural networks meet personalized pagerank. In: International Conference on Learning Representations (ICLR), New Orleans, 2019.

[42] Bai S J, Zico Kolter J, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv. Available: http://arxiv.org/abs/1803.01271. Accessed on [2023-11-10].

[43] Jiang B, Zhang Z, Lin D D, Tang J, Luo B. Semi-supervised learning with graph learning-convolutional networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach: IEEE, 2019; pp.11313–11320. doi: 10.1109/CVPR.2019.01157.

[44] Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2018; 32: 11118105.

[45] Chen F W, Pan S R, Jiang J, Huo H, Long G D. DAGCN: Dual attention graph convolutional networks. In 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, 2019; pp.1–8. doi: 10.1109/IJCNN.2019.8851698.

[46] Abu-El-Haija S, Perozzi B, Kapoor A, Alipourfard N, Lerman K, Harutyunyan H, et al. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In: Proceedings of the Proceedings of the 36th International Conference on Machine Learning, 2019; pp.21–29.

[47] Lim B, Zohren S. Time-series forecasting with deep learning: A survey. Philosophical Transactions of the Royal Society A-mathematical Physical and Engi, 2021; 379(2194): 20200209.

[48] Chen H L, Yang J B, Fu X H, Zheng Q X, Song X Y, Fu Z D, et al. Water quality prediction based on LSTM and attention mechanism: A case study of the Burnett River, Australia. Sustainability, 2022; 14(20): 13231.

[49] Zuo C H, Wang J L, Liu M P, Deng S H, Wang Q N. An ensemble framework for short-term load forecasting based on TimesNet and TCN. Energies, 2023; 16(14): 5330.

[50] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS' 16). Curran Associates Inc., Red Hook, 2016; pp.3844–3852.

[51] Gehring J, Auli M, Grangier D, Yarats D, Dauphin Y N. Convolutional sequence to sequence learning. In: Proceedings of the Proceedings of the 34th International Conference on Machine Learning, PMLR, 2017; pp.1243–1252.

[52] Chen H H, Zhu M Y, Hu X, Wang J R, Sun Y, Yang J D, et al. Multifeature short-term power load forecasting based on GCN-LSTM. International Transactions on Electrical Energy Systems, 2023; 2023(1): 8846554.

[53] Du S D, Li T R, Yang Y, Horng S J. Multivariate time series forecasting via attention-based encoder-decoder framework. Neurocomputing, 2020; 388: 269–279.

[54] Liu G, Jiang Y, Zhong K, Yang Y, Wang Y. A time series model adapted to multiple environments for recirculating aquaculture systems. Aquaculture, 2023; 567: 739284.

[55] Mohammadi H A, Ghofrani S, Nikseresht A. Using empirical wavelet transform and high-order fuzzy cognitive maps for time series forecasting. Applied Soft Computing, 2023; 135: 109990.