Environmental Science and Ecotechnology 23 (2025) 100514

Contents lists available at ScienceDirect

Environmental Science and Ecotechnology

journal homepage: www.journals.elsevier.com/environmental-science-and-ecotechnology/

Original Research

Graph neural networks and transfer entropy enhance forecasting of mesozooplankton community dynamics

Minhyuk Jeung [a], Min-Chul Jang [b, **], Kyoungsoon Shin [b], Seung Won Jung [c], Sang-Soo Baek [d, *]

[a] Department of Rural & Biosystems Engineering (Brain Korea 21), Chonnam National University, Gwangju, 61186, Republic of Korea
[b] Ballast Water Research Center, Korea Institute of Ocean Science & Technology, Geoje, 53201, Republic of Korea
[c] Library of Marine Samples, Korea Institute of Ocean Science & Technology, Geoje, 53201, Republic of Korea
[d] Department of Environmental Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea

ARTICLE INFO

Article history:
Received 12 June 2024
Received in revised form
19 November 2024
Accepted 19 November 2024

Keywords:
Graph neural network
Ecosystem dynamics
Mesozooplankton
Transfer entropy

ABSTRACT

Mesozooplankton are critical components of marine ecosystems, acting as key intermediaries between primary producers and higher trophic levels by grazing on phytoplankton and influencing fish populations. They play pivotal roles in the pelagic food web and export production, affecting the biogeochemical cycling of carbon and nutrients. Therefore, accurately modeling and visualizing mesozooplankton community dynamics is essential for understanding marine ecosystem patterns and informing effective management strategies. However, modeling these dynamics remains challenging due to the complex interplay among physical, chemical, and biological factors, and the detailed parameterization and feedback mechanisms are not fully understood in theory-driven models. Graph neural network (GNN) models offer a promising approach to forecast multivariate features and define correlations among input variables. The high interpretive power of GNNs provides deep insights into the structural relationships among variables, serving as a connection matrix in deep learning algorithms. However, there is insufficient understanding of how interactions between input variables affect model outputs during training. Here we investigate how the graph structure of ecosystem dynamics used to train GNN models affects their forecasting accuracy for mesozooplankton species. We find that forecasting accuracy is closely related to interactions within ecosystem dynamics. Notably, increasing the number of nodes does not always enhance model performance; closely connected species tend to produce similar forecasting outputs in terms of trend and peak timing. Therefore, we demonstrate that incorporating the graph structure of ecosystem dynamics can improve the accuracy of mesozooplankton modeling by providing influential information about species of interest. These findings will provide insights into the influential factors affecting mesozooplankton species and emphasize the importance of constructing appropriate graphs for forecasting these species.

© 2024 The Authors. Published by Elsevier B.V. on behalf of Chinese Society for Environmental Sciences, Harbin Institute of Technology, Chinese Research Academy of Environmental Sciences. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Mesozooplankton species are important components of marine ecosystems, where they act as grazers of phytoplankton and influencers of fish populations, which is an essential pathway for transferring energy between primary producers and organisms at higher trophic levels [1–3]. The dynamics of the mesozooplankton community, including factors such as growth, mortality, and distribution, shape the structure of the ecosystems and are also used to evaluate the impact of global change [4,5]. They also play a role in marine biogeochemical cycles by converting organic matter into dissolved inorganic carbon and nutrient pools [6]. They are also responsible for transmitting energy to deep water through diel vertical migration and can play a significant role in deoxygenating the upper ocean [7,8]. Mesozooplankton communities are highly sensitive to oceanographic conditions, with their life cycles and

* Corresponding author.
** Corresponding author.
E-mail addresses: mcjang@kiost.ac.kr (M.-C. Jang), ssbaek@yu.ac.kr (S.-S. Baek).

https://doi.org/10.1016/j.ese.2024.100514
2666-4984/© 2024 The Authors. Published by Elsevier B.V. on behalf of Chinese Society for Environmental Sciences, Harbin Institute of Technology, Chinese Research Academy of Environmental Sciences. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

List of Abbreviations

Adam	Adaptive moment estimation
ANN	Artificial neural networks
Chl-T	Total chlorophyll
DFT	Discrete Fourier transform
FC	Fully connected
GFT	Graph Fourier transform
GNN	Graph neural networks
GRU	Gated recurrent unit
KIOST	Korea Institute of Ocean Science and Technology
LSTM	Long-short-term memory
NSE	Nash–Sutcliffe efficiency
PBIAS	Percent bias
RNN	Recurrent neural network
StemGNN	Spectral-temporal graph neural network
TE	Transfer entropy

community responses particularly affected by environmental changes [9]. Therefore, understanding and modeling the temporal changes in mesozooplankton abundance and dynamics are crucial for recognizing patterns in marine ecosystems and developing effective management approaches [10,11].

Modeling and interpreting mesozooplankton dynamics are challenging due to their direct and sensitive responses to various physical, chemical, and biological changes within ecosystems [12,13]. Therefore, the models must adequately capture the dynamics of these interactions. The modeling of zooplankton is commonly classified into biological, ecosystem, and size-based models [10], which are extremely sensitive to the parameterization of zooplankton, wherein poor parameterization adversely affects model performance [14,15]. Previous research has significantly advanced the mechanistic modeling of zooplankton by improving the representation of zooplankton functional parameters [16]. Nevertheless, these models often group different zooplankton species into the same functional category or distinguish them based solely on size [17], making it difficult to understand the individual characteristics and temporal changes of specific mesozooplankton. To address this issue, individual-based models have been utilized to simulate the behavior and physiology of specific zooplankton species [18,19]. These models require high computational costs and extensive species-specific information for parameterization and are primarily applied to well-investigated species [10].

Deep learning methods, such as long-short-term memory (LSTM), artificial neural networks (ANN), and regression transformer models using multivariate input data [20–22], have been applied to improve forecasting accuracy and reduce computational costs in modeling zooplankton species. These models differ from theory-driven models, which compute zooplankton dynamics using parameters or feedback with variables. Instead, they utilize data-based information, such as temporal dependencies and weights from other influential features [23]. Despite their success, these methods have several limitations. Previous studies have often trained models to simulate single-target zooplankton species separately, increasing the number of models with increased target species [24]. Furthermore, there has been insufficient consideration of how interactions between input variables affect model outputs during training [25]. Simulating multivariate time series data is challenging when relationships between input variables are complex, and the number of target variables increases [26]. To overcome these limitations, methods that consider the relationships between input variables to improve model accuracy and interpretability are receiving more attention [27].

The multivariate algae or community forecasting approach can provide a perspective on biological–environmental interactions, which enables the identification of the interactive effects of multiple input variables on model responses [28]. The graph theory is a method to generate the relationships between each variable with several nodes and edges graphically. This method has been used in a wide range of applications, such as defining the relationships in social networks [29,30], financial market [31], and transportation [32,33]. Due to its high interpretive power, the graph neural network (GNN) approach has been used in various applications with complex interactions between variables, as it can effectively capture spatial relationships [34]. GNNs are ideally suited for traffic forecasting problems [35], communication networks [36], and disease transmission [37], which can be represented using non-Euclidean graph structures. The traditional GNN models are typically based on fixed (or static) graph structures and cannot capture temporal evolution [38], rendering them unsuitable for capturing temporal dynamics in ecosystems. Therefore, advanced GNN models have been developed to consider the correlation among multivariate data that exhibit temporal variations [34]. Despite advances in GNN models, most of them require a dependency graph as a prior, which is difficult to apply in fields where it is challenging to fully define the interactions or relationships between variables, such as biological fields. Cao et al. proposed the spectral-temporal graph neural network (StemGNN) that performs better by automatically capturing temporal patterns and correlations in the spectral domain [39]. To the best of our knowledge, limited research uses GNN models for predicting mesozooplankton community dynamics and interpreting the impact of various meteorological and environmental factors on forecasting results.

Hence, in this study, we aimed to relate the forecasting performance of mesozooplankton communities to the graph structure of ecosystem dynamics to understand how variables affect forecasting accuracy. We applied the StemGNN model at various lead times and with different graph structure scenarios to (1) determine how far into the future accurate predictions can be made and (2) evaluate the association between prediction accuracy and input features consisting of graph structures. Moreover, we compared the StemGNN model's effectiveness in considering inter-series relationships and temporal dependencies with a developed LSTM model. We also visualized ecosystem dynamics in graph structures for the entire period and seasonally to explore temporal changes in factor influence. Finally, based on our findings, we discuss the results and limitations and suggest future directions for forecasting mesozooplankton community dynamics.

2. Materials and methods

2.1. Study site and data acquisition

The Jinhae Bay is a semi-enclosed embayment located on the Geoje Island, southeastern coast of the Republic of Korea. The sampling site (34°59'41" N, 128°40'31" E) was selected near the laboratory (South Sea Research Institute, Korea Institute of Ocean Science and Technology (KIOST); approximately 300 m from the sampling site) to acquire *in situ* water samples at a weekly interval (Fig. S1). Sampling was conducted at a depth of 8–10 m, with a tide range of approximately 2.2 m, during high tide (±1 h). Triplicate vertical tows were performed around high tide using a conical 45-cm-diameter net with 200-μm mesh to sample mesozooplankton. Each mesozooplankton species was then counted using Bogorov's counting chamber under a microscope. Groups of Dinoflagellata, Cnidaria, Chordata, Chaetognatha, Arthropoda, Copepoda, Meroplankton, and Ichthyoplankton (a total of 30 mesozooplankton species) were investigated through a long-term monitoring program of KIOST [40], which recorded data over ten years from

January 2001 to December 2020. The environmental (e.g., water temperature, salinity, pH, chlorophyll, and seven other water quality constituents) and meteorological (e.g., minimum–maximum temperature, rainfall, wind speed, humidity, and sun duration) factors were also monitored to identify the dynamics of mesozooplankton communities under various environmental conditions. We collected 995 data samples of 30 mesozooplankton taxa, 11 environmental factors, and 6 meteorological factors (Supplementary Materials Tables S1 and S2).

2.2. The overall procedure for modeling and interpretation of mesozooplankton community dynamics using GNNs

The dynamics of mesozooplankton communities were forecasted by incorporating interactions between ecosystem factors. Modeling and interpreting mesozooplankton dynamics improve our understanding of linkages among trophic levels and help examine potential ecosystem mechanisms [5,10]. Mesozooplankton interact with various nutrients, and the nutrient composition required for their metabolic growth differs between species [41]. This complexity limits our understanding of mesozooplankton growth processes and reduces model accuracy [42]. The GNN method is a powerful deep-learning technique for forecasting complex, interconnected time series data by leveraging their interactions [43]. Providing an appropriate graph structure is crucial for determining model performance [44], whereas inaccurately defining the relationships between input variables can adversely impact predictive outcomes [45]. To overcome these limitations, it is essential to apply deep learning methods that automatically capture the interactions among input variables.

The StemGNN model (Fig. 1a), which automatically captures the correlations between input variables, was applied using the Python software (detailed explanations in Section 2.3.1). As mentioned, forecasting mesozooplankton community dynamics has been limited despite using other deep learning methods such as ANN and LSTM. These models are inadequate for capturing spatial relationships (i.e., correlations between input features) in the forecasting process [46]. Therefore, the LSTM model was used as a baseline to compare the performances of the StemGNN model and to evaluate the effectiveness of considering spatial dependencies in the StemGNN model. Although the StemGNN model achieves excellent performance by capturing the dynamic correlations between input variables, it may present a challenge in interpreting the model outputs [34]. The interpretation of ecosystem dynamics can provide rich cognitive insights into the structure of mutually beneficial relationships and negative correlations such as predation, consumption, and competition [47]. To explore the interactions within ecosystems, we applied the transfer entropy (TE) theory for computing nonlinear relationships from time series data of algal species, water quality constituents, and meteorological factors (Fig. 1b). TE calculation and graph visualization were performed using the graph and network algorithms functions and ProcessNetwork application in the MATLAB software [48] (detailed explanations in the Supplementary Materials).

2.3. Modeling mesozooplankton community dynamics using the multivariate forecast method

Multiple mesozooplankton species were simultaneously forecasted by leveraging the interactions between features rather than forecasting each species separately. Multivariate forecasting methods utilize dependencies between input features to improve model performance [49]. This approach addresses complex decision-making problems [50,51] and evaluates ecosystem trade-offs and synergies [52]. Among multivariate models, GNNs have

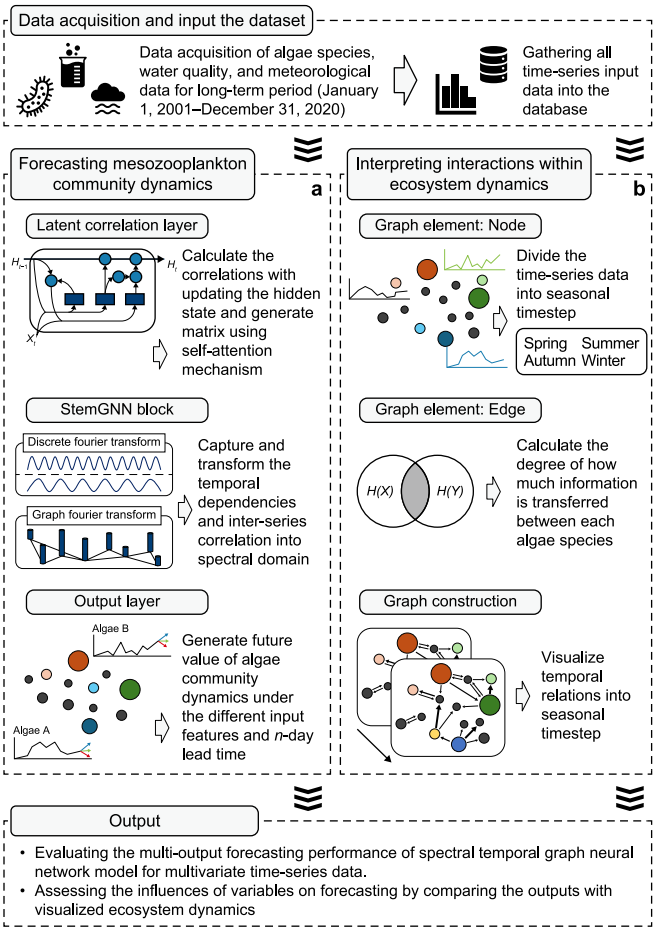

Fig. 1. Overall procedure for forecasting mesozooplankton community dynamics using spectral-temporal graph neural network (StemGNN) and visualizing ecosystem dynamics using transfer entropy.

gained attention for their ability to exploit relationships between multiple time series inputs [53]. This study applied a GNN model that utilizes the correlations between variables to forecast mesozooplankton communities (Fig. 1a).

2.3.1. The StemGNN model

The StemGNN was applied as a forecasting model due to its powerful ability to forecast multivariate time series data [54]. This method provides better forecasting performance by jointly capturing the inter-series correlations and temporal dependencies [39]. Other current state-of-the-art models also incorporate spatial and temporal dependencies in traffic forecasting and communication network prediction; however, most require a dependency graph as a prior [39]. The key technical contribution of the StemGNN model is its ability to automatically compute correlations in multivariate time series data without depending on predefined topologies, thereby improving its applicability across a wide range of applications. The StemGNN model applies a global prediction mechanism, forecasting at the level of the entire graph by considering information from all nodes and edges within the graph (i.e., the input and output dimensions must be equal). Despite the advantages of enabling the model to understand and leverage interactions between variables, a limitation is that it cannot select target variables (i.e., all input variables are forecasted and fed into a sliding window, and a rolling strategy is used to forecast the next timestep). Nevertheless, it is inadequate to forecast meteorological

and water quality variables using interactions from mesozooplankton species. Before training the model, we modified the source code of the StemGNN model to ensure that the predicted values of mesozooplankton species and the observed data for meteorological and chemical factors were fed into the sliding window before forecasting the next timestep.

The dataset was partitioned into three subsets for training, validation, and testing with a ratio of 6:3:1. The parameter settings were optimized manually as follows: 100 epochs, 14 window size, 64 batch size, and 0.001 learning rate with a 0.7 decay rate after every five epochs. Forecasting outputs with longer lead times (or forecasting horizons) are crucial for providing response time to address algal blooms [55], and the StemGNN model was applied across different forecasting horizons (or lead times), viz., 1, 7, and 14 days. The StemGNN model was constructed with three layers: the latent correlation layer, the StemGNN block, and the output layer. The overall architecture of the StemGNN method is processed as follows:

(1) The correlations of multivariate time series data in the latent correlation layer are calculated (Supplementary Material Fig. S2a). In this layer, the multivariate input data are fed into the gated recurrent unit (GRU) model that updates the hidden state in each timestamp t sequentially. When the last hidden state was updated, the adjacency matrix for constructing the graph structure was calculated by the self-attention mechanism as equation (1):

$$Q = RW^Q, K = RW^K, W = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \quad (1)$$

where Q and K are the query and key of the inputted weight matrix W. The output matrix $W \in \mathbb{R}^{N\times N}$ is transformed as an adjacency matrix for the graph structure.

(2) The constructed graph structure is transferred into the StemGNN block (Supplementary Material Fig. S2b), designed to capture the temporal dependencies of multivariate time series data and joint with inter-series correlation information (i.e., the output of the latent correlation layer). Then, these two information are transformed into a spectral domain using graph Fourier transform (GFT) and discrete Fourier transform (DFT). Cao et al. have described more detailed information regarding the StemGNN block [39].
(3) After calculating the inter-series correlations and temporal dependencies using GFT and DFT operators, the output is supplied into the graph convolution layer, and inverse GFT is performed. The fully connected (FC) sublayers comprise two types of output layers. The forecasting layer generates the future values, whereas the backcasting layer improves multivariate time series data representation power.
(4) The forecasted value is fed into the sliding window, and a rolling strategy is used for multistep forecasting. This process is repeated for each forecasting horizon (or lead time) set until the target date we intend to predict.

2.3.2. Model evaluations

The LSTM model, which can capture the temporal dependency of input features [22], was developed and used as a baseline for comparing the performances of the StemGNN model and demonstrating the effectiveness of capturing correlations between input variables in the forecasting process of StemGNN. The critical difference between the StemGNN and LSTM models is that the StemGNN model can consider both temporal dependencies and inter-series correlations [39]. In contrast, the LSTM model considers only the temporal dependency of input features [56]. The LSTM model is formulated from the recurrent neural network (RNN) architecture. The LSTM network consists of an input layer, an LSTM layer, a dropout layer, and an output layer. The closed-loop method was applied to the LSTM model using the predicted value of the previous time step as input data. However, like the StemGNN model, only the predicted values of mesozooplankton species were used for the next step of forecasting, and the observed data for meteorological and chemical factors were fed into a sliding window before forecasting the next timestep. The parameter settings for the LSTM layer were optimized using the adaptive moment estimation (Adam) optimizer with 2000 epochs, a batch size of 64, and a learning rate of 0.001. The Adam optimizer is an advanced version compared with classical stochastic gradient descent methods and is known for its fast convergence speed [57].

To evaluate the forecasting models (i.e., LSTM and StemGNN), the Nash–Sutcliffe efficiency (NSE) and the percent bias (PBIAS) were used as the quantitative objective functions, both of which are widely used as evaluation criteria in modeling studies [58]. The NSE is useful for interpreting scale-varying datasets because it determines the relative magnitude of the residual variance compared with the measured data variance [59]. The PBIAS also provides information on whether the simulated data are larger or smaller than the observed data and is useful for long-term simulations [60]. Both quantitative objective functions were calculated using the following equations:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \quad (2)$$

$$PBIAS\ (\%) = \frac{\sum_{i=1}^{n}(X_i - Y_i) \times 100}{\sum_{i=1}^{n} X_i} \quad (3)$$

where X_i and Y_i are the observed and predicted values, and n is the number of samples.

2.4. Evaluating model predictions according to different graph structure scenarios

The effect of input variables within the graph structure on the predicted output was evaluated using the StemGNN model by combining the graph components. The composition of graphs (i.e., the number of nodes and the correlation between nodes) significantly impacts model performance [61]. It provides insights into the contributions of input variables to the prediction outputs [62]. The five graph structure scenarios described in Table 1 were derived from two objectives, i.e., to evaluate the impact of the number of nodes and to identify the effects of the predictive power of nodes. Determining the number of nodes is important in preventing model degradation caused by high dimensionality [63]. Scenarios I and II were structured with single and multiple mesozooplankton while retaining environmental factors in the graph structure. In these scenarios, the performances of the LSTM and StemGNN models were compared to determine the effectiveness of capturing the correlations between input variables in the forecasting process of the StemGNN model. Scenarios III–V were composed by combining the five mesozooplankton species with

Table 1
Input features and objectives for five different scenarios.

Graph structure scenarios	Objectives	Components of graph structure
Scenario I	Assessing the impact of the number of nodes on model performance	Meteorological data, water quality data, single target mesozooplankton species[a]
Scenario II		Meteorological data, water quality data, entire mesozooplankton species
Scenario III	Identifying the impact of the predictive power of nodes on model performance	Meteorological data, water quality data, five mesozooplankton species with the highest accuracy
Scenario IV		Meteorological data, water quality data, five mesozooplankton species with the lowest accuracy
Scenario V		Meteorological data, water quality data, five mesozooplankton species with the highest accuracy and five mesozooplankton species with the lowest accuracy

[a] 30 species were individually inputted for prediction.

the highest accuracy (i.e., well-predicted mesozooplankton species), the five mesozooplankton species with the lowest accuracy (i.e., poorly predicted mesozooplankton species), and the ten mesozooplankton species with both the highest and lowest accuracy. This was performed by ranking the forecasting performance of each mesozooplankton species in Scenario I.

3. Results

3.1. Weather records and monitoring data

The meteorological factors and water quality constituents were collected as node variables of graph structures and the StemGNN model. Descriptive statistics of the weather and monitoring data are provided in Table S1 (Supplementary Material). The daily meteorological data, comprising 7300 records, and the weekly environmental data, consisting of 995 observations (e.g., mesozooplankton species and water quality), were collected from January 4, 2001, to December 29, 2020. The Jinhae Bay has four seasons, and the dates for the beginning and end of each season are as follows: spring runs from March 1 to May 31, summer is from June 1 to August 31, autumn runs from September 1 to November 30, and winter runs from December 1 to February 28. The study site receives an average annual precipitation of 873 mm and 45% in summer but only 149 mm and 8% in winter.

The summary statistics of mesozooplankton species are described in Table S2 and Fig. S3 (Supplementary Materials). The mesozooplankton communities were primarily dominated by groups of Dinoflagellata (*Noctiluca scintillans)*, Arthropoda (especially *Penilia avirostris*), and Copepoda (especially *Paracalanus parvus*, *Acartia omorii*, and Copepodites) (red box in Supplementary Material Fig. S3a). *N. scintillans*, *P. parvus*, and Copepodites maintained an abundance rate of >90% during all seasons (Supplementary Material Fig. S3b), implying that these species have a relatively low sensitivity to fluctuations in meteorological conditions, such as substantial changes in rainfall and temperature over the seasons.

3.2. Forecasting mesozooplankton community dynamics using GNNs

3.2.1. Forecasting results based on different graph structures

A comparative analysis evaluated the prediction accuracy of 30 different mesozooplankton species based on different graph structure scenarios (Table 1). The StemGNN model accurately predicted mesozooplankton species in Scenarios I and II, with average NSE values of 0.833 and 0.869, respectively. However, the predictive performances for *Calanus sinicus* and *Centropages abdominalis* were excluded from the evaluation due to their poor performance (negative NSE values; Fig. 2 and Supplementary Material Tables S3–S4). The graph obtained using the entire mesozooplankton (Scenario II) produced more stable results, with NSE values > 0.8 for most species, whereas Scenario I outcomes were underestimated and distinctly separated into groups with NSE values < 0.8 and > 0.9. However, regarding the LSTM model, there was a significant variation in performance across different mesozooplankton species, with Scenarios I and II showing average NSE values of 0.479 and 0.411 (without *Calanus sinicus* and *Centropages abdominalis*), respectively (Supplementary Material Table S5). Compared with the LSTM model, the StemGNN model demonstrated performance improvements ranging from 73.9% (in Scenario I) to 111.8% (in Scenario II) in terms of average NSE values (Fig. 3). Although the StemGNN demonstrated improvements in Scenario II, there was no significant improvement in the LSTM model, implying that the LSTM model cannot effectively utilize the interactions between mesozooplankton species during the forecasting process.

After forecasting multiple mesozooplankton species using the graph structures of Scenarios I and II, the five mesozooplankton species with the highest accuracy and the five mesozooplankton species with the lowest accuracy were classified to construct the graph structures for Scenarios III–V (Table 1). Forecasting with Scenarios III–V aimed to identify the impact of the predictive power of the nodes on model performance, determining whether the model performs better or worse when the graph is constructed from well-predicted or poorly predicted species. According to the model performance on Scenario I, the five well-predicted (cirripede larvae, *Sagitta crassa*, *Penilia avirostris*, *Noctiluca scintillans*, and fish eggs) and five poorly predicted (*Calanus sinicus*, *Centropages abdominalis*, *Podon* spp., *Pseudodiaptomus marinus*, and tunicate larvae) species were classified by ranking the NSE values (bolded species in Supplementary Material Table S3).

The forecasting results for the five well-predicted species closely matched the observed values in both low and high abundances, with an average NSE of 0.974 (Fig. 4), indicating that the StemGNN model accurately captured both the temporal trends and peak abundance timings of these species. These results emphasize the efficacy of the latent correlation layer within the StemGNN model, demonstrating remarkable performance without prior knowledge of the ecosystem's interactions. Conversely, the poorly predicted species (five of the worst-predicted plankton in Scenario II) exhibited significant discrepancies, with an average NSE value of −0.725, including both underestimations and overestimations (Fig. 5). Among these species, adequate predictive performance was observed for *Pseudodiaptomus marinus*, *Podon* spp., and tunicate larvae in capturing the peak timing, whereas the peak abundance was underestimated. However, for *Calanus sinicus* and *Centropages abdominalis*, the model significantly failed to capture both trend and peak abundance timing. The model output for *Centropages abdominalis* in July 2020 emulated the trend for *Calanus sinicus* (Fig. 5), indicating the possibility of mutual effects when two nodes are connected in the graph structure.

Comparisons were made across the five graph structure scenarios (Table 1) to investigate the impact of input variable numbers

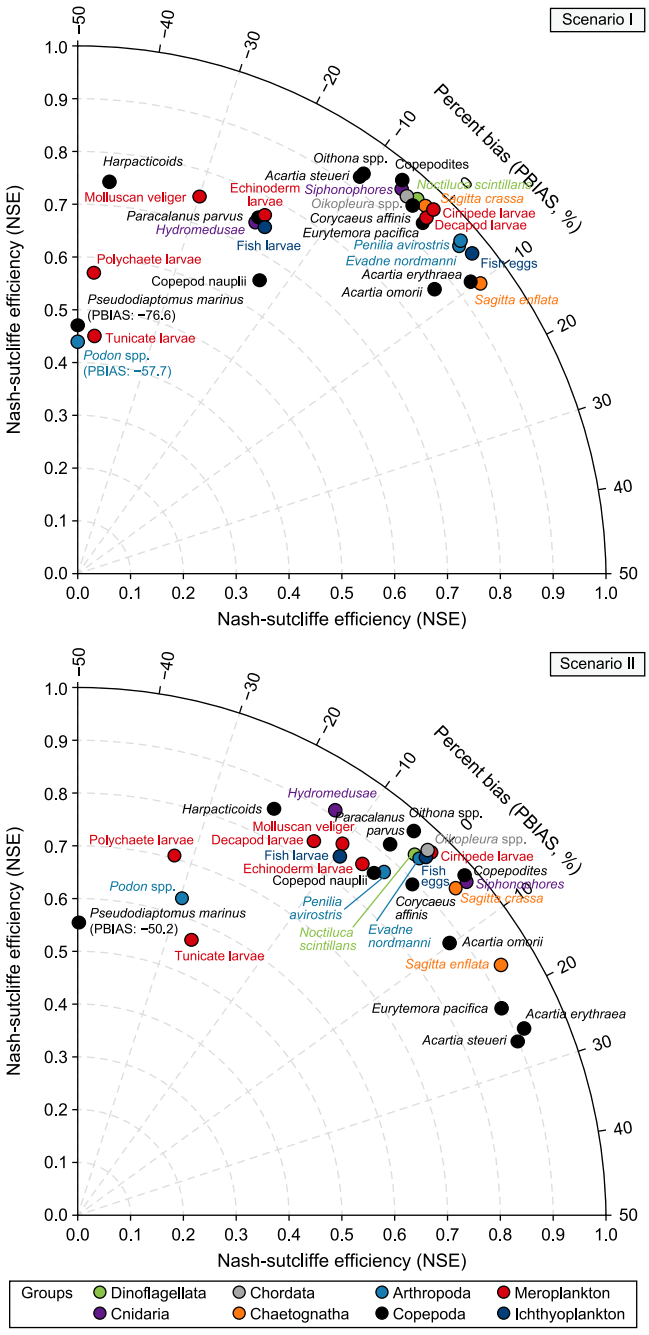

Fig. 2. Forecasting performance of the 1-day lead time spectral-temporal graph neural network (StemGNN) model for 30 mesozooplankton species (Scenarios I and II) during the test period (01/01/2019–12/29/2020). The species *Calanus sinicus* and *Centropages abdominalis* were excluded due to low Nash–Sutcliffe efficiency (NSE) values that exceeded the scale of the figure.

and their predictive capacities on model performance (Fig. 6). For well-predicted species, no significant difference was observed across input feature scenarios, with the coefficient of variance averaging 1.9%. This result suggests that these species are weakly connected or disconnected from poorly forecasted species or offset negative influences due to their strong relationships with meteorological and water quality factors. Among the poorly predicted species, the forecasting performance for the group of *Calanus sinicus* and *Centropages abdominalis* and a group of *Podon* spp., *Pseudodiaptomus marinus*, and tunicate larvae exhibited contrasting

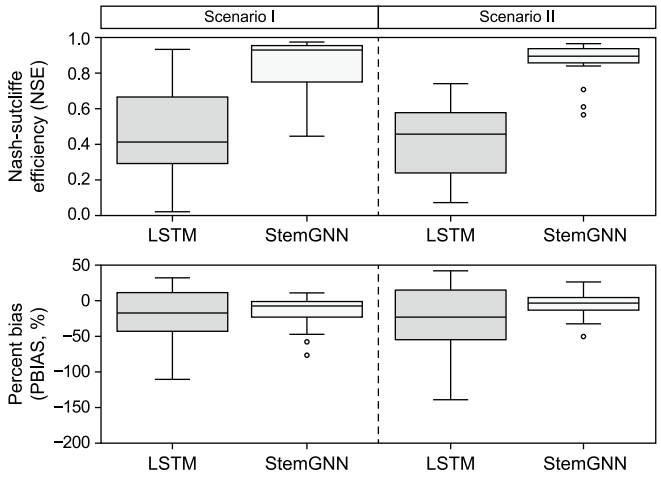

Fig. 3. Comparison of forecasting accuracy (Nash–Sutcliffe efficiency, NSE) and percent bias (PBIAS) of the 1-day lead time long-short-term memory (LSTM) model and spectral-temporal graph neural network (StemGNN) model for 30 mesozooplankton species (Scenarios I and II) during the test period (January 1, 2019–December 29, 2020). The species *Calanus sinicus* and *Centropages abdominalis* were excluded due to poor NSE and PBIAS values that exceeded the scale of the figure.

results (Fig. 6b). The forecasting performances for *Calanus sinicus* and *Centropages abdominalis* did not improve when well-predicted species and other mesozooplankton species were applied in the graph (i.e., negative NSE value for the five graph structure scenarios). Forecasting for the group of *Podon* spp., *Pseudodiaptomus marinus*, and tunicate larvae was initially underestimated but showed improved performance when either well-predicted species or the entire mesozooplankton species (i.e., Scenarios II and V) were applied.

3.2.2. Comparison of forecasting performance across different lead times

The model performance across different lead times, including peak timing and temporal trends, was compared with the prediction results of Scenario II across 1-, 7-, and 14-day lead times. In terms of NSE, forecasting performances consistently improved when the lead times were shortened from 14 to 1 day (Fig. 7). The 1-day forecasting horizon (blue line in Fig. 7) provided an average NSE value of 0.869 in the prediction of mesozooplankton species. In contrast, the 7- and 14-day lead times showed average NSE values of 0.788 and 0.723, respectively. For PBIAS, as expected, the 1-day prediction demonstrated the most reliable performance (i.e., close to the blue line, which implied an unbiased estimation). However, the PBIAS of 7- and 14-day lead time exhibited inconsistent results. Forecasting for *Pseudodiaptomus marinus* was underestimated in the 7-day lead time (PBIAS: −146.848%) compared with the 14-day lead time (PBIAS: −60.545%).

3.3. Seasonal changes in interactions within ecosystem dynamics and lag time

3.3.1. Seasonal changes in ecological variable interactions

The interaction strength within ecosystem dynamics was visualized by the length and width of edges for the entire period and each season (Fig. 8; Supplementary Material Fig. S6). The graph structure divided mesozooplankton species from meteorological and water quality data for the entire period. Specifically, sunlight duration, rainfall, and total chlorophyll (Chl-T) were identified as significant factors, each with multiple connections to mesozooplankton species (12, 13, and 11 connections, respectively). No

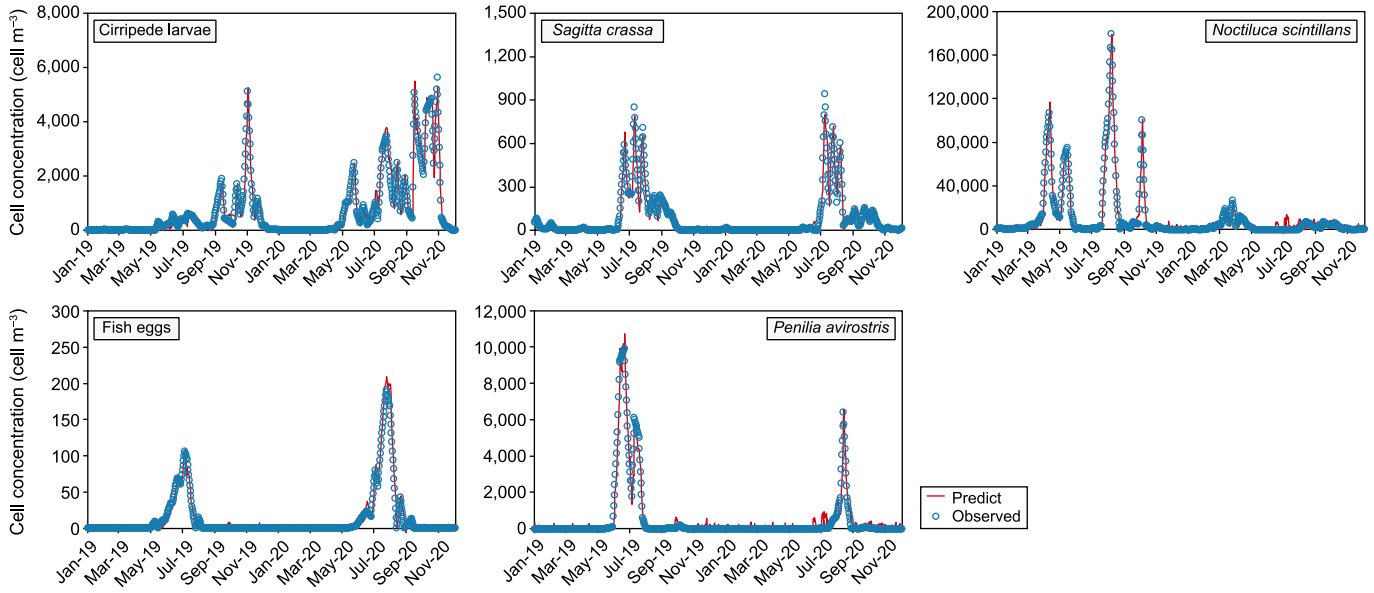

Fig. 4. Forecasting results of Scenario I (i.e., environmental factors and single mesozooplankton species were applied as input features) for five mesozooplankton species with the highest accuracy with 1-day lead time during the test period (January 1, 2019–December 29, 2020).

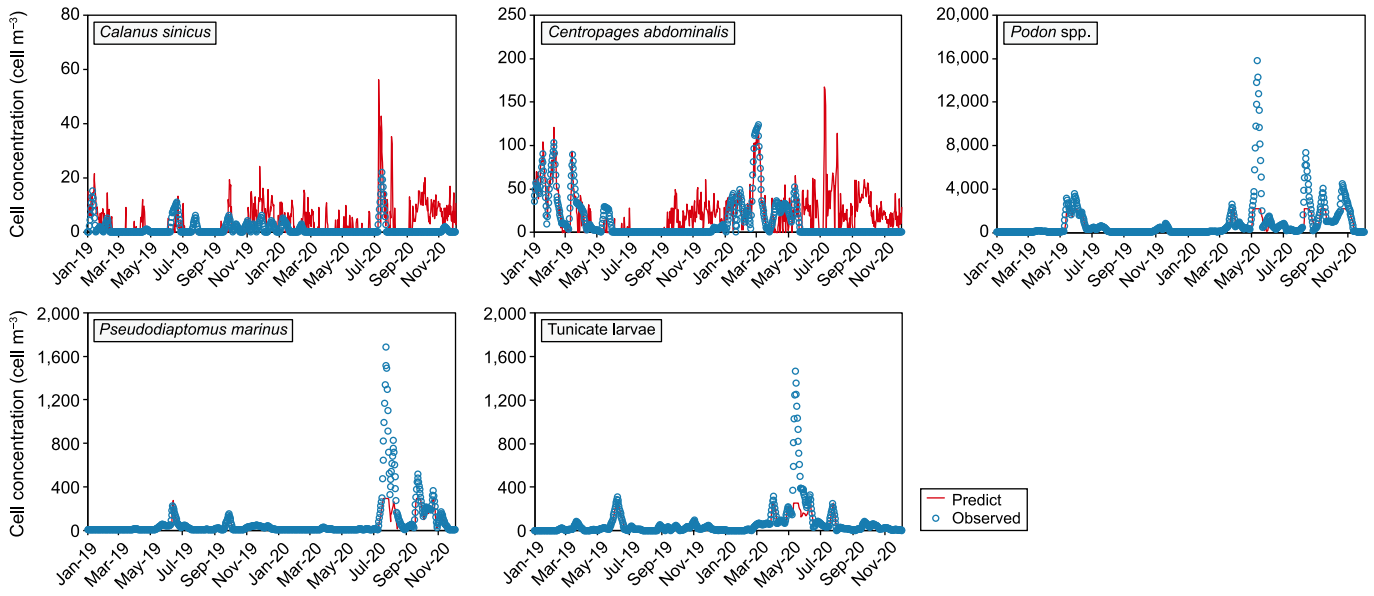

Fig. 5. Forecasting results of Scenario I (i.e., environmental factors and single mesozooplankton species were applied as input features) for five mesozooplankton species with the lowest accuracy with 1-day lead time during the test period (January 1, 2019–December 29, 2020).

specific variables were crucial across all four seasons in the seasonal graph structure. Rainfall and sun duration significantly impacted mesozooplankton species in autumn, summer, and winter, whereas the effects of meteorological factors decreased in spring (Supplementary Material Fig. S6). This result suggests that the impact of environmental conditions on ecological dynamics varies seasonally.

Thus, we confirmed the variance of influences between factors under seasonal changes (Fig. 9). This variance was quantified by the increase (blue bar) or decrease (red bar) in TE values between variables as seasons changed. Meteorological factors demonstrated significant seasonal variation, fluctuating by an average of 0.045 bits between seasons, whereas the influence between plankton species was smaller, fluctuating by 0.015 bits on average. Remarkably, the influence of sun duration on various species increased by 0.073 bits on average when transitioning from summer to autumn (blue-colored bar in Fig. 9a). Moreover, the influence of sun duration decreased (0.082 bits on average) when the season changed from autumn to winter, whereas temperature became a more influential variable in winter (increment of 0.079 bits on average; blue-colored bar in Fig. 9b).

3.3.2. Lag times between ecological variables

The lag time between each variable was determined by applying lag times ranging from 1 to 30 days and identifying the maximum TE value (Supplementary Material equation S(3)). A black bar

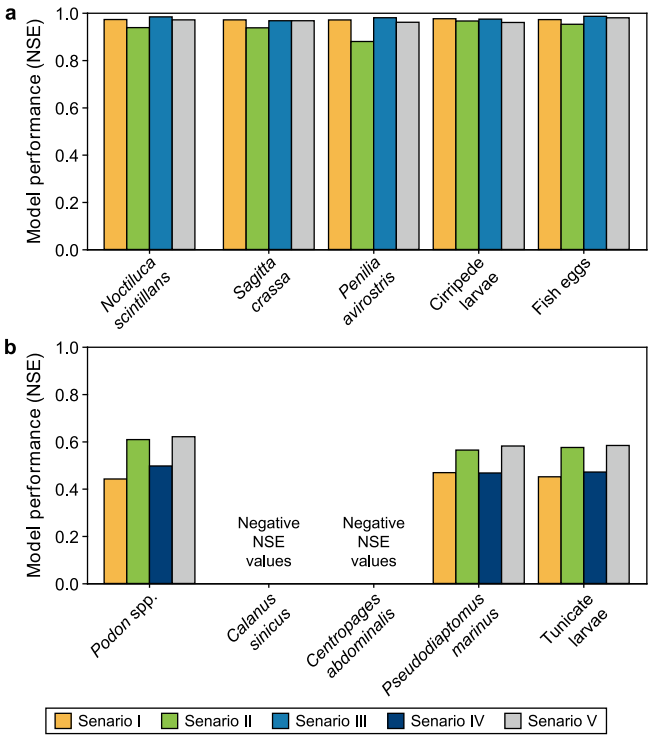

Fig. 6. Forecasting performance of the spectral-temporal graph neural network (StemGNN) model with 1-day lead time during the test period (January 1, 2019–December 29, 2020) under different input feature scenarios (input feature scenarios from I to V). **a**, Five species with the highest accuracy; **b**, Five species with the lowest accuracy.

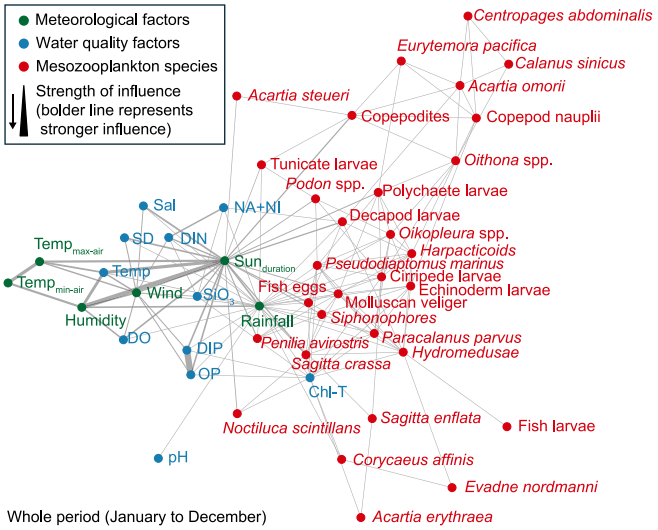

Fig. 8. The constructed graph structure in the mesozooplankton species group, water quality, and meteorological factors (nodes) for the entire monitoring period (January 4, 2001–December 29, 2020). The width and length of the line (edges) between the nodes imply the strength of influences. The meteorological factors include minimum and maximum air temperature ($Temp_{air}$), wind speed (Wind), relative humidity (Humidity), sunshine duration ($Sun_{duration}$), and rainfall. The water quality factors include dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP), organic phosphorus (OP), salinity (Sal), water temperature (Temp), dissolved oxygen (DO), Secchi disk depth (SD), total chlorophyll (Chl-T), silicate (SiO_3), pH, and sodium plus nickel (NA + NI).

represented the transfer of information between each variable in a day. Furthermore, a larger portion of the black bar indicates strong influences within a day. We compared the differences in lag time between species by coloring maximized TE from 1 to 30 days from blue to red. Meteorological factors influenced most mesozooplankton species within a shorter period (e.g., >80% of the maximum TE value was black; upper part of Supplementary Materials Figs. S7–S9). However, water quality factors require a medium-to-long delay to influence species and interactions between mesozooplankton (e.g., the colored bar often exceeds 50% of

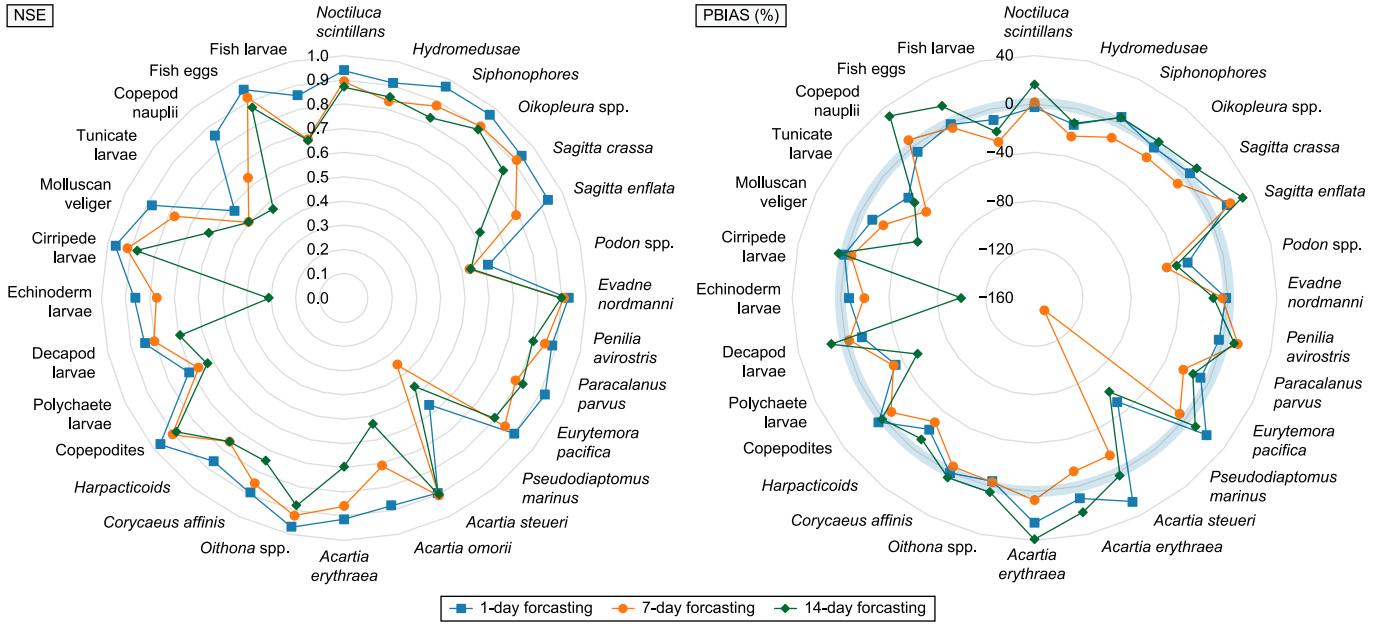

Fig. 7. Comparison of forecasting accuracy (Nash–Sutcliffe efficiency, NSE) and percent bias (PBIAS) according to different lead times (1, 7, and 14 days) in Scenario II. The blue transparent circle for PBIAS represents the best performance. The species *Calanus sinicus* and *Centropages abdominalis* were excluded due to low performance, which degraded the visibility of the figure.

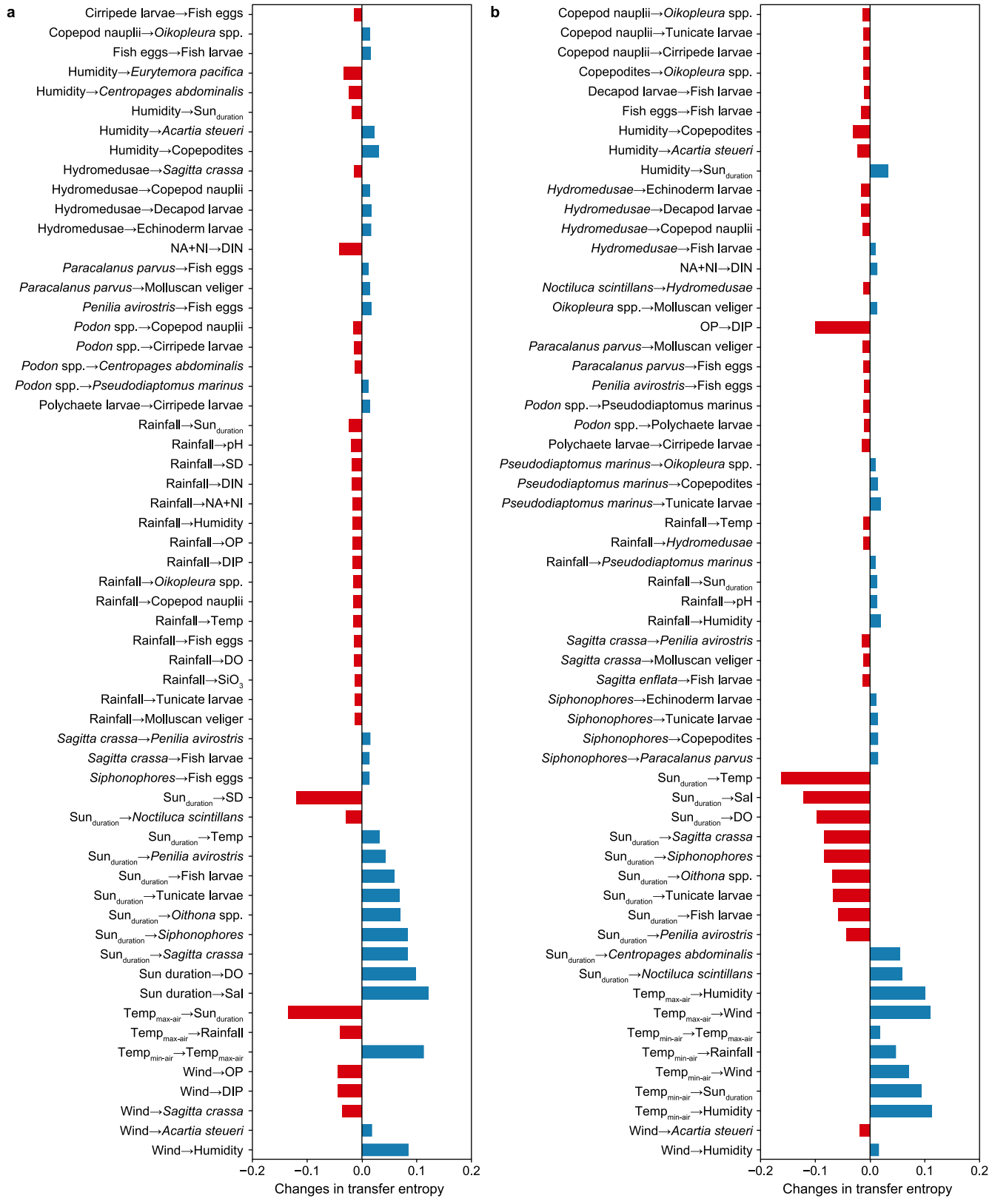

Fig. 9. Seasonal changes in influences (transfer entropy) between each variable: **a**, from summer to autumn; **b**, from autumn to winter. The values in the interquartile range (i.e., slight changes) are excluded. The negative value (red bar) represents a decrease of influences, whereas the positive value (blue bar) implies an increase of influences between two variables. The meteorological factors include minimum and maximum air temperature ($Temp_{air}$), wind speed (Wind), relative humidity (Humidity), sunshine duration ($Sun_{duration}$), and rainfall. The water quality factors include dissolved inorganic nitrogen (DIN), dissolved inorganic phosphorus (DIP), organic phosphorus (OP), salinity (Sal), water temperature (Temp), dissolved oxygen (DO), Secchi disk depth (SD), total chlorophyll (Chl-T), silicate (SiO_3), pH, and sodium plus nickel (NA + NI).

the portion; see the lower part of Supplementary Materials Figs. S7–S9).

For the five well-predicted mesozooplankton species (cirripede larvae, *Sagitta crassa*, *Penilia avirostris*, *Noctiluca scintillans*, and fish eggs), the TE of most variables was maximized with blue or green bars (Supplementary Material Fig. S7). The difference in TE values between black and color bars was smaller than that in the poorly predicted species (i.e., instantaneous influences from other variables to well-predicted species). This result indicates that most variables effectively influence these five mesozooplankton species in less than 15 days. The TE values for the five poorly predicted mesozooplankton species (*Calanus sinicus*, *Centropages abdominalis*, *Podon* spp., *Pseudodiaptomus marinus*, and tunicate larvae) were maximized with longer lag times, ranging from yellow to red bar (Supplementary Material Fig. S8). The TE values of meteorological factors (e.g., sun duration and humidity) were maximized over 20 days.

4. Discussion

4.1. Understanding temporal distribution and interactions within ecosystem dynamics

We used graph structures to visually represent seasonal changes to analyze ecosystem dynamics. Ecosystem dynamics are formed through interactions between meteorological, water quality, and other organisms [64], and these variables play crucial roles in shaping species abundance or richness [65]. Among the mesozooplankton species, *Noctiluca scintillans* and Copepodites were dominant in both abundance and richness (10,708 cells m^{-3} and 96% abundance rate during the entire period; Supplementary Material Fig. S3) and may be influential factors for other species. Copepodites were associated with six different mesozooplankton species, whereas *Noctiluca scintillans* demonstrated a relationship with one species. When species related to Copepodites were considered, most species belonged to the same taxonomic groups. This suggests that the relationships between mesozooplankton species are limited within certain groups due to group-specific temporal differences in life strategies [66]. These findings support the need to monitor other biological variables (e.g., phytoplankton and fish) to consider the biological relationships in the GNN model.

The significance of environmental factors in shaping the abundance and distribution of mesozooplankton species is widely recognized [9]. Remarkably, the duration of sunlight and rainfall strongly influences these species. We assume that connections exist between these factors and mesozooplankton species. Rainfall drives nitrogen and phosphorus loading, increasing nutrient abundance [67,68]. Sunlight duration influences phytoplankton growth by providing energy for photosynthesis [69]. Among the water quality factors, Chl-T is widely related to mesozooplankton species (connected with 11 mesozooplankton species). The impact of Chl-T is well-known, as nutrient loading may initially elevate the food availability for zooplankton [70]. Salinity and water temperature, which are significant determinants [71], showed no direct correlation with other mesozooplankton species. Studies have revealed a major shift in zooplankton dynamics occurs when certain salinity or water temperature thresholds are reached [72,73]. Moreover, increased salinity can result in cascading trophic interactions, indicating that these factors indirectly affect zooplankton dynamics through upper or lower trophic levels [74].

4.2. Comparison of forecasting accuracy between the LSTM and StemGNN models

The StemGNN model demonstrated significant improvements compared with the LSTM model (Fig. 3), which considers only the temporal dependencies of input variables by memorizing time series information. This result suggests that forecasting mesozooplankton community dynamics is highly complex and that the StemGNN model has an advantage in utilizing the interactions between individual mesozooplankton species to improve model outputs. Cao et al. also confirmed that the GFT layer, a component of StemGNN that captures inter-series relationships among input variables, effectively improves forecasting performance by leveraging the inter-series information within a graph [39]. Moreover, Lu et al. revealed that GNN models can improve performance by capturing the interactions between input variables and utilizing latent properties to model temporal and correlation dependencies [46]. In this context, forecasting complex, interconnected variables using graph structures can improve accuracy, making the selection of appropriate input variables an important process before training the StemGNN model, which may affect forecasting accuracy.

4.3. Comparison of forecasting accuracy under different graph structure scenarios

The performances of forecasting accuracy varied according to the graph structure scenarios (Table 1), which emphasized the impact of interconnected nodes or the number of input features on the model outputs. By comparing Scenario I (wherein environmental factors and each mesozooplankton species were applied) and Scenario II (wherein environmental factors and 30 mesozooplankton species were applied), we identified a double-edged sword effect of forecasting performance between well-predicted and poorly predicted mesozooplankton species (Fig. 2). In Scenario II, the model's forecasting accuracy for the poorly predicted mesozooplankton species (i.e., tunicate larvae, *Pseudodiaptomus marinus*, *Podon* spp., polychaete larvae, and Copepod nauplii) remained insufficient; however, there was a significant improvement in terms of prediction accuracies (e.g., 27% increment of NSE on average). Conversely, for *Noctiluca scintillans*, *Sagitta crassa*, *Penilia avirostris*, fish eggs, and cirripede larvae, the forecasting performance degraded (e.g., a 4% decrement of NSE on average) when the number of mesozooplankton species was included in the graph structure (i.e., Scenario II). These data suggest that increasing the number of connections between mesozooplankton species in the graph structure does not always improve performance, depending on the target variable. This may be due to increased sensitivity to unpredictable variables or outliers when calculating the total loss function [75,76]. Maurya et al. argued that using more nodes can result in acquiring meaningless information that does not improve performance [77]. Furthermore, only mesozooplankton was considered a biological variable in our study. We assume adding variables with upper or lower trophic interactions (e.g., phytoplankton and fish) would improve the performance.

The output of the StemGNN model reveals that correlated nodes exhibit temporal variation and peak timing similarities. For instance, *Calanus sinicus* and *Centropages abdominalis* exhibited similar trends and peak timing despite having different temporal patterns in observed values, which resulted in worse forecasting performance (Fig. 5). We also confirmed that *Calanus sinicus* and *Centropages abdominalis* were closely located within the graph structure for the entire period (Fig. 8). This finding indicates that closer nodes may strongly influence together, irrespective of temporal patterns or peak timing. This finding is consistent with a previous study that reported closer nodes indicated similar results or performance on GNN applications. Qi et al. emphasized that substantial differences in the abundance or diversity of neighboring nodes can cause uncertainty [78].

4.4. Comparison of forecasting accuracy across different lead times

A comparison of forecasting accuracy across different lead times can provide insights into how far into the future the model can make accurate predictions [79]. Based on the results of forecasting accuracies under different lead times (1, 7, and 14 days), the StemGNN model consistently provided more accurate predictions when the lead time was shortened (Fig. 7). In terms of NSE, the model's prediction accuracy degraded from 9.4% to 16.8% on average when lead times increased from 1 to 7 and 14 days (the predictive performances for *Calanus sinicus* and *Centropages abdominalis* were excluded due to their poor performance). Although the model's accuracy decreased, 7- and 14-day forecasts provided adequate accuracy, with the NSE value being >0.8 for 13 mesozooplankton species across the 14-day lead time. This result indicates good agreement with observed values regarding trend and peak timing for several mesozooplankton species. Therefore, these findings confirm the ability of the GNN model to predict mesozooplankton dynamics and its potential for bloom detection.

4.5. Further research suggestions

Although the GNN model demonstrated significant performance in forecasting mesozooplankton community dynamics, several issues must be addressed. First, the performance of the GNN model should be evaluated separately within essential or related target variables. This study examined model performance variations when different nodes were applied to the GNN model. According to the results, larger target variables may degrade the model performance when calculating the loss function. For instance, increasing the number of target variables can increase the sensitivity to unpredictable variables (i.e., variables that exhibit seriously poor performance) or outliers in forecast results when calculating the total loss function [75,76]. Second, incorporating additional factors from a broader perspective (e.g., the ocean in our experiment) into the graph may improve the accuracy of models. For instance, variables such as phytoplankton and predators are well-known, influential factors of mesozooplankton [1,80], which can be considered predictor variables for forecasting mesozooplankton. However, we did not consider these factors and used only Chl-T as a proxy for phytoplankton biomass. Third, the lag time between each variable should be considered when training the GNN model. According to the aquatic food chains, biological delay systems have been of considerable interest for a long time [81]. The abundance of zooplankton species decreases after a specific lag time of toxic phytoplankton bloom [82]. Therefore, applying the lag time to input variables can improve the model's performance by improving the correlation in specific timing between input and output variables [83]. Fourth, utilizing GNN explanation techniques, such as GNNExplainer [84], may facilitate the comprehension of key graph pathways and pivotal node feature information. This study investigated the significance of interaction between each node in predicting mesozooplankton dynamics based on the graph structure. Nevertheless, interpreting the graph structure provides information only about whether each node is associated, thus making it difficult to distinguish positive and negative associations [85]. Finally, monitoring a broader range of mesozooplankton species could significantly improve the generalization capabilities of the model and facilitate more meaningful comparisons between different models. The previous model was specifically developed, optimized, and evaluated for a limited set of monitored species. Nonetheless, when applied to other species that are not included in the training set of the model, the performance of the model may deteriorate, and comparison between models becomes unfeasible [42]. Consequently, it is important to develop the model for a diverse array of species to improve its generalization capability and enable the evaluation of improvements in future model developments.

5. Conclusions

Based on the modeling experiment, this study revealed that the forecasting accuracy for each mesozooplankton species is closely associated with interactions in the ecosystem dynamics of historical data. Mesozooplankton dynamics were simulated using the StemGNN model, and then, the graph network was visualized to identify the relationship between forecasting performance. Regarding the closely connected mesozooplankton species, the forecasting output was also influenced regarding trend and peak timing, indicating that significant differences among nearby nodes can introduce uncertainty in forecasting. Increasing the number of nodes does not always benefit the model performance, which improves sensitivity to unpredictable variables or outliers when calculating the total loss. The forecasting performance for mesozooplankton species was maximized with a 1-day lead time; however, the accuracy decreased by 9.4–16.8% for the 7- and 14-day lead times. Despite this decrease, the forecasting output for a 14-day lead time provides good agreement regarding trend and peak timing. Incorporating additional input features from a broad perspective, such as grazers and predators, and considering biological delay times in the GNN model should be addressed in future work to improve the accuracy of models. These results clarify the relationship between forecasting performance and interactions among variables in ecosystem dynamics, emphasizing the importance of constructing appropriate graphs for forecasting mesozooplankton species.

CRediT authorship contribution statement

Minhyuk Jeung: Writing - Original Draft, Validation, Software, Formal Analysis, Conceptualization, Writing - Review & Editing. **Min-Chul Jang:** Writing - Review & Editing, Supervision, Methodology. **Kyoungsoon Shin:** Validation, Formal Analysis, Data Curation. **Seung Won Jung:** Writing - Review & Editing, Visualization, Data Curation. **Sang-Soo Baek:** Writing - Review & Editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Ministry of Science and ICT (MSIT) through a Sejong Science Fellowship funded by the National Research Foundation of Korea (NRF) (No. 2022R1C1C2003649).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ese.2024.100514.

References

[1] M. Chen, H. Liu, S. Song, J. Sun, Size-fractionated mesozooplankton biomass and grazing impact on phytoplankton in northern South China Sea during four seasons, Deep Sea Res. Part II Top. Stud. Oceanogr. 117 (2015) 108–118.
[2] L.H. Sweat, H. Alexander, E.J. Phlips, K.B. Johnson, Mesozooplankton community dynamics and grazing potential across algal bloom cycles in a

subtropical estuary, Front. Mar. Sci. 8 (2021) 734270.
[3] R.F. Heneghan, J.D. Everett, J.L. Blanchard, A.J. Richardson, Zooplankton are not fish: improving zooplankton realism in size-spectrum models mediates energy transfer in food webs, Front. Mar. Sci. 3 (2016).
[4] J.E. Keister, D. Bonnet, S. Chiba, C.L. Johnson, D.L. Mackas, R. Escribano, Zooplankton population connections, community dynamics, and climate variability, ICES (Int. Counc. Explor. Sea) J. Mar. Sci. 69 (2012) 347–350.
[5] H. Liu, M.J. Fogarty, J.A. Hare, C.-h. Hsieh, S.M. Glaser, H. Ye, E. Deyle, G. Sugihara, Modeling dynamic interactions and coherence between marine zooplankton and fishes linked to environmental variability, J. Mar. Syst. 131 (2014) 120–129.
[6] C. Clerc, O. Aumont, L. Bopp, Should we account for mesozooplankton reproduction and ontogenetic growth in biogeochemical modeling? Theor. Ecol. 14 (2021) 589–609.
[7] A. Ariza, J.C. Garijo, J.M. Landeira, F. Bordes, S. Hernández-León, Migrant biomass and respiratory carbon flux by zooplankton and micronekton in the subtropical northeast Atlantic Ocean (Canary Islands), Prog. Oceanogr. 134 (2015) 330–342.
[8] D. Bianchi, E.D. Galbraith, D.A. Carozza, K.A.S. Mislan, C.A. Stock, Intensification of open-ocean oxygen depletion by vertically migrating animals, Nat. Geosci. 6 (2013) 545–548.
[9] D.L. Mackas, G. Beaugrand, Comparisons of zooplankton time series, J. Mar. Syst. 79 (2010) 286–304.
[10] J.D. Everett, M.E. Baird, P. Buchanan, C. Bulman, C. Davies, R. Downie, C. Griffiths, R. Heneghan, R.J. Kloser, L. Laiolo, A. Lara-Lopez, H. Lozano-Montes, R.J. Matear, F. McEnnulty, B. Robson, W. Rochester, J. Skerratt, J.A. Smith, J. Strzelecki, I.M. Suthers, K.M. Swadling, P. van Ruth, A.J. Richardson, Modeling what we sample and sampling what we model: challenges for zooplankton model assessment, Front. Mar. Sci. 4 (2017).
[11] E. Wong, C. Yau, K.Y.K. Chan, Seasonal and spatial dynamics of mesozooplankton community in a subtropical embayment, Regional Studies in Marine Science 56 (2022) 102724.
[12] S.C. Marques, U.M. Azeiteiro, S.M. Leandro, H. Queiroga, A.L. Primo, F. Martinho, I. Viegas, M.Â. Pardal, Predicting zooplankton response to environmental changes in a temperate estuarine ecosystem, Mar. Biol. 155 (2008) 531–541.
[13] L. Sousa, K.O. Coyle, R.P. Barry, T.J. Weingartner, R.R. Hopcroft, Climate-related variability in abundance of mesozooplankton in the northern Gulf of Alaska 1998–2009, Deep Sea Res. Part II Top. Stud. Oceanogr. 132 (2016) 122–135.
[14] F. Carlotti, J.C. Poggiale, Towards methodological approaches to implement the zooplankton component in “end to end” food-web models, Prog. Oceanogr. 84 (2010) 20–38.
[15] A. Mitra, Are closure terms appropriate or necessary descriptors of zooplankton loss in nutrient–phytoplankton–zooplankton type models? Ecol. Model. 220 (2009) 611–620.
[16] C. Le Quéré, E.T. Buitenhuis, R. Moriarty, S. Alvain, O. Aumont, L. Bopp, S. Chollet, C. Enright, D.J. Franklin, R.J. Geider, S.P. Harrison, A.G. Hirst, S. Larsen, L. Legendre, T. Platt, I.C. Prentice, R.B. Rivkin, S. Sailley, S. Sathyendranath, N. Stephens, M. Vogt, S.M. Vallina, Role of zooplankton dynamics for Southern Ocean phytoplankton biomass and global biogeochemical cycles, Biogeosciences 13 (2016) 4111–4133.
[17] E. Nogueira, J.D. Woods, C. Harris, A.J. Field, S. Talbot, Phytoplankton coexistence: results from an individual-based simulation model, Ecol. Model. 198 (2006) 1–22.
[18] R. Bi, H. Liu, Effects of variability among individuals on zooplankton population dynamics under environmental conditions, Mar. Ecol. Prog. Ser. 564 (2017) 9–28.
[19] G. Perhar, N.E. Kelly, F.J. Ni, M.J. Simpson, A.J. Simpson, G.B. Arhonditsis, Using Daphnia physiology to drive food web dynamics: a theoretical revisit of Lotka-Volterra models, Ecol. Inf. 35 (2016) 29–42.
[20] S.-S. Baek, E.-Y. Jung, J. Pyo, Y. Pachepsky, H. Son, K.H. Cho, Hierarchical deep learning model to simulate phytoplankton at phylum/class and genus levels and zooplankton at the genus level, Water Res. 218 (2022) 118494.
[21] A. Banerjee, M. Chakrabarty, N. Rakshit, A.R. Bhowmick, S. Ray, Environmental factors as indicators of dissolved oxygen concentration and zooplankton abundance: deep learning versus traditional regression approach, Ecol. Indicat. 100 (2019) 99–117.
[22] M. Liu, J. He, Y. Huang, T. Tang, J. Hu, X. Xiao, Algal bloom forecasting with time-frequency analysis: a hybrid deep learning approach, Water Res. 219 (2022) 118591.
[23] M.M. Taye, Understanding of machine learning with deep learning: architectures, workflow, applications and future directions, Computers 12 (2023) 91.
[24] N. Kim, J. Shin, Y. Cha, Multisite algal bloom predictions in a lake using graph attention networks, Environmental Engineering Research 29 (2024) 230210, 230210.
[25] T. Kim, J. Shin, D. Lee, Y. Kim, E. Na, J.-h. Park, C. Lim, Y. Cha, Simultaneous feature engineering and interpretation: forecasting harmful algal blooms using a deep learning approach, Water Res. 215 (2022) 118289.
[26] A. Kouzuma, K. Watanabe, Exploring the potential of algae/bacteria interactions, Curr. Opin. Biotechnol. 33 (2015) 125–129.
[27] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: a review of methods and applications, AI Open 1 (2020) 57–81.
[28] D.B. Botkin, H. Saxe, M.B. Araújo, R. Betts, R.H.W. Bradshaw, T. Cedhagen, P. Chesson, T.P. Dawson, J.R. Etterson, D.P. Faith, S. Ferrier, A. Guisan, A.S. Hansen, D.W. Hilbert, C. Loehle, C. Margules, M. New, M.J. Sobel, D.R.B. Stockwell, Forecasting the effects of global warming on biodiversity, Bioscience 57 (2007) 227–236.
[29] A. Alamsyah, B. Rahardjo, Social Network Analysis Taxonomy Based on Graph Representation, 2021 arXiv preprint arXiv:2102.08888.
[30] S.M. Kostić, M.I. Simić, M.V. Kostić, Social network analysis and churn prediction in telecommunications using graph theory, Entropy 22 (2020) 753.
[31] D. Cheng, F. Yang, S. Xiang, J. Liu, Financial time series forecasting with multi-modality graph neural network, Pattern Recogn. 121 (2022) 108218.
[32] Z. Hu, F. Shao, R. Sun, A new perspective on traffic flow prediction: a graph spatial-temporal network with complex network information, Electronics 11 (2022) 2432.
[33] W. Liu, Y. Zheng, S. Chawla, J. Yuan, X. Xing, Discovering spatio-temporal causal interactions in traffic data streams, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, San Diego, California, USA, 2011, pp. 1010–1018.
[34] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu, A comprehensive survey on graph neural networks, IEEE Transact. Neural Networks Learn. Syst. 32 (2021) 4–24.
[35] W. Jiang, J. Luo, Graph neural network for traffic forecasting: a survey, Expert Syst. Appl. 207 (2022) 117921.
[36] W. Jiang, Graph-based deep learning for communication networks: a survey, Comput. Commun. 185 (2022) 40–54.
[37] F.H.O. Duarte, G.J.P. Moreira, E.J.S. Luz, L.B.L. Santos, V.L.S. Freitas, Time series forecasting of COVID-19 cases in Brazil with GNN and mobility networks, in: M.C. Naldi, R.A.C. Bianchi (Eds.), Intelligent Systems, Springer, Nature Switzerland, Cham, 2023, pp. 361–375.
[38] A. Longa, V. Lachi, G. Santin, M. Bianchini, B. Lepri, P. Lio, F. Scarselli, A. Passerini, Graph Neural Networks for Temporal Graphs: State of the Art, Open Challenges, and Opportunities, 2023 arXiv preprint arXiv:2302.01018.
[39] D. Cao, Y. Wang, J. Duan, C. Zhang, X. Zhu, C. Huang, Y. Tong, B. Xu, J. Bai, J. Tong, Spectral temporal graph neural network for multivariate time-series forecasting, Adv. Neural Inf. Process. Syst. 33 (2020) 17766–17778.
[40] M.C. Jang, P.G. Jang, H.G. Cha, W.J. Lee, M.K. Bae, J.S. Kang, B. Hyun, K.S. Shin, Long-term trends in mesozooplankton community at a coastal station in Jinhae Bay from 2001 to 2020, in: Proceedings of the 2023 KAOSTS Conference, 2023, p. 111.
[41] E. Acheampong, M.H. Nielsen, A. Mitra, M.A. St John, Towards an adaptive model for simulating growth of marine mesozooplankton: a macromolecular perspective, Ecol. Model. 225 (2012) 1–18.
[42] L. Ratnarajah, R. Abu-Alhaija, A. Atkinson, S. Batten, N.J. Bax, K.S. Bernard, G. Canonico, A. Cornils, J.D. Everett, M. Grigoratou, N.H.A. Ishak, D. Johns, F. Lombard, E. Muxagata, C. Ostle, S. Pitois, A.J. Richardson, K. Schmidt, L. Stemmann, K.M. Swadling, G. Yang, L. Yebra, Monitoring and modelling marine zooplankton in a changing climate, Nat. Commun. 14 (2023) 564.
[43] Q. Zhang, J. Chang, G. Meng, S. Xu, S. Xiang, C. Pan, Learning graph structure via graph convolutional networks, Pattern Recogn. 95 (2019) 308–318.
[44] Y. Shi, P. Quan, Y. Xiao, M. Lei, L. Niu, Graph influence network, IEEE Trans. Cybern. 53 (2023) 6146–6159.
[45] Z. Jin, Y. Wang, Q. Wang, Y. Ming, T. Ma, H. Qu, GNNLens: a visual analytics approach for prediction error diagnosis of graph neural networks, IEEE Trans. Visual. Comput. Graph. 29 (2023) 3024–3038.
[46] Z. Lu, W. Lv, Y. Cao, Z. Xie, H. Peng, B. Du, LSTM variants meet graph neural networks for road speed prediction, Neurocomputing 400 (2020) 34–45.
[47] M. Kruk, E. Paturej, Indices of trophic and competitive relations in a planktonic network of a shallow, temperate lagoon. A graph and structural equation modeling approach, Ecol. Indicat. 112 (2020) 106007.
[48] B.L. Ruddell, C. Sturtevant, M. Kang, R. Yu, ProcessNetwork Software, 2008, version 1.5. https://github.com/ProcessNetwork/ProcessNetwork_Software.
[49] V.G. Satorras, S.S. Rangapuram, T. Januschowski, Multivariate Time Series Forecasting with Latent Graph Inference, 2022 arXiv preprint arXiv:2203.03423.
[50] E. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, I. Vlahavas, Multi-target regression via input space expansion: treating targets as inputs, Mach. Learn. 104 (2016) 55–98.
[51] D. Xu, Y. Shi, I.W. Tsang, Y.S. Ong, C. Gong, X. Shen, Survey on multi-output learning, IEEE Transact. Neural Networks Learn. Syst. 31 (2020) 2409–2429.
[52] J.L. Ticehurst, L.T.H. Newham, D. Rissik, R.A. Letcher, A.J. Jakeman, A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia, Environ. Model. Software 22 (2007) 1129–1139.
[53] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, Adv. Neural Inf. Process. Syst. 29 (2016).
[54] S. Wang, Y. Zhang, E. Fu, S. Tang, Multiscale backcast convolution neural network for traffic flow prediction in the frequency domain, Appl. Sci. 12 (2022) 11912.
[55] W.M. Woelmer, R.Q. Thomas, M.E. Lofton, R.P. McClure, H.L. Wander, C.C. Carey, Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability, Ecol. Appl. 32 (7) (2022) 1–22.
[56] S. Yang, S. Zhong, K. Chen, W-WaveNet: a multi-site water quality prediction model incorporating adaptive graph convolution and CNN-LSTM, PLoS One 19 (2024) e0276155.

[57] A. Kumar, S. Sarkar, C. Pradhan, Malaria disease detection using CNN technique with SGD, RMSprop and ADAM optimizers, in: S. Dash, B.R. Acharya, M. Mittal, A. Abraham, A. Kelemen (Eds.), Deep Learning Techniques for Biomedical and Health Informatics, Springer International Publishing, Cham, 2020, pp. 211–230.
[58] H.V. Gupta, H. Kling, K.K. Yilmaz, G.F. Martinez, Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling, J. Hydrol. 377 (2009) 80–91.
[59] J.E. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models part I—a discussion of principles, J. Hydrol. 10 (1970) 282–290.
[60] D.N. Moriasi, M.W. Gitau, N. Pai, P. Daggupati, Hydrologic and water quality models: performance measures and evaluation criteria, Transactions of the ASABE 58 (2015) 1763–1785.
[61] L. Liu, R.-C. Chen, A novel passenger flow prediction model using deep learning methods, Transport. Res. C Emerg. Technol. 84 (2017) 74–91.
[62] N. Liu, Q. Feng, X. Hu, Interpretability in Graph Neural Networks. Graph Neural Networks: Foundations, frontiers, and applications, 2022, pp. 121–147.
[63] D.B. Acharya, H. Zhang, Feature selection and extraction for graph neural networks, in: Proceedings of the 2020 ACM Southeast Conference, 2020, pp. 252–255.
[64] Y. Gao, Q. Yang, H. Li, X. Wang, A. Zhan, Anthropogenic pollutant-driven geographical distribution of mesozooplankton communities in estuarine areas of the Bohai Sea, China, Sci. Rep. 9 (2019) 9668.
[65] T. Kodama, T. Wagawa, N. Iguchi, Y. Takada, T. Takahashi, K.I. Fukudome, H. Morimoto, T. Goto, Spatial variations in zooplankton community structure along the Japanese coastline in the Japan Sea: influence of the coastal current, Ocean Sci. 14 (2018) 355–369.
[66] R. Klais, M. Lehtiniemi, G. Rubene, A. Semenova, P. Margonski, A. Ikauniece, M. Simm, A. Põllumäe, E. Grinienė, K. Mäkinen, H. Ojaveer, Spatial and temporal variability of zooplankton in a temperate semi-enclosed sea: implications for monitoring design and long-term studies, J. Plankton Res. 38 (2016) 652–661.
[67] T.C. Ballard, E. Sinha, A.M. Michalak, Long-term changes in precipitation and temperature have already impacted nitrogen loading, Environ. Sci. Technol. 53 (2019) 5080–5090.
[68] E. Sinha, A.M. Michalak, Precipitation dominates interannual variability of riverine nitrogen loading across the continental United States, Environ. Sci. Technol. 50 (2016) 12874–12884.
[69] Q. Zhou, W. Chen, K. Shan, L. Zheng, L. Song, Influence of sunlight on the proliferation of cyanobacterial blooms and its potential applications in Lake Taihu, China, J. Environ. Sci. 26 (2014) 626–635.
[70] M. Chen, B. Chen, P. Harrison, H. Liu, Dynamics of mesozooplankton assemblages in subtropical coastal waters of Hong Kong: a comparative study between a eutrophic estuarine and a mesotrophic coastal site, Continent. Shelf Res. 31 (2011) 1075–1086.
[71] P. Ezhilarasan, C.K. Basuri, A. Gera, M. Kumaraswami, V.R. Rao, M.V.R. Murthy, Mesozooplankton distribution in relation to the salinity gradient in a tropical hypersaline lake, J. Sea Res. 178 (2021) 102138.
[72] Z. Horváth, C.F. Vad, A. Tóth, K. Zsuga, E. Boros, L. Vörös, R. Ptacnik, Opposing patterns of zooplankton diversity and functioning along a natural stress gradient: when the going gets tough, the tough get going, Oikos 123 (2014) 461–471.
[73] Q. Lin, L. Xu, J. Hou, Z. Liu, E. Jeppesen, B.-P. Han, Responses of trophic structure and zooplankton community to salinity and temperature in Tibetan lakes: implication for the effect of climate warming, Water Res. 124 (2017) 618–629.
[74] Z. Ersoy, M. Abril, M. Cañedo-Argüelles, C. Espinosa, L. Vendrell-Puigmitja, L. Proia, Experimental assessment of salinization effects on freshwater zooplankton communities and their trophic interactions under eutrophic conditions, Environ. Pollut. 313 (2022) 120127.
[75] W. Chao, L. Bo, W. Lei, P. Pai, Improving boosting methods with a stable loss function handling outliers, International Journal of Machine Learning and Cybernetics 14 (2023) 2333–2352.
[76] I. Razzak, K. Zafar, M. Imran, G. Xu, Randomized nonlinear one-class support vector machines with bounded loss function to detect of outliers for large scale IoT data, Future Generat. Comput. Syst. 112 (2020) 715–723.
[77] S.K. Maurya, X. Liu, T. Murata, Improving Graph Neural Networks with Simple Architecture Design, 2021 arXiv preprint arXiv:2105.07634.
[78] Y. Qi, Q. Li, H. Karimian, D. Liu, A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory, Sci. Total Environ. 664 (2019) 1–10.
[79] O.L. Petchey, M. Pontarp, T.M. Massie, S. Kéfi, A. Ozgul, M. Weilenmann, G.M. Palamara, F. Altermatt, B. Matthews, J.M. Levine, D.Z. Childs, B.J. McGill, M.E. Schaepman, B. Schmid, P. Spaak, A.P. Beckerman, F. Pennekamp, I.S. Pearse, The Ecological Forecast Horizon, and examples of its uses and determinants, bioRxiv (2015) 013441.
[80] L.C. Stige, P. Dalpadado, E. Orlova, A.-C. Boulay, J.M. Durant, G. Ottersen, N.C. Stenseth, Spatiotemporal statistical analyses reveal predator-driven zooplankton fluctuations in the Barents Sea, Prog. Oceanogr. 120 (2014) 243–253.
[81] R.R. Sarkar, B. Mukhopadhyay, R. Bhattacharyya, S. Banerjee, Time lags can control algal bloom in two harmful phytoplankton–zooplankton system, Appl. Math. Comput. 186 (2007) 445–459.
[82] J. Chattopadhyay, R.R. Sarkar, A.e. Abdllaoui, A delay differential equation model on harmful algal blooms in the presence of toxic substances. Mathematical Medicine and Biology, A Journal of the IMA 19 (2002) 137–161.
[83] X. Chen, J. Huang, Z. Han, H. Gao, M. Liu, Z. Li, X. Liu, Q. Li, H. Qi, Y. Huang, The importance of short lag-time in the runoff forecasting model based on long short-term memory, J. Hydrol. 589 (2020) 125359.
[84] Z. Ying, D. Bourgeois, J. You, M. Zitnik, J. Leskovec, Gnnexplainer: generating explanations for graph neural networks, Adv. Neural Inf. Process. Syst. 32 (2019).
[85] J.T. Siddons, A.J. Irwin, Z.V. Finkel, Graphical analysis of A marine plankton community reveals spatial, temporal, and niche structure of sub-communities, Front. Mar. Sci. 9 (2022).