Journal of *Marine Science and Engineering*

MDPI

Article

Multi-Step Forecasting of Chlorophyll Concentration with Multi-Attention Collaborative Network

Yingying Jin [1], Feng Zhang [1,*], Xia Wang [2], Lei Wang [3], Kuo Chen [1], Liangyu Chen [1], Yutao Qin [4] and Ping Wu [3]

[1] East Sea Information Center, State Oceanic Administration, Shanghai 200136, China; jinyingying@ecs.mnr.gov.cn (Y.J.); chenkuo@ecs.mnr.gov.cn (K.C.); athena_cly@126.com (L.C.)
[2] School of Teacher Education, Shangqiu Normal University, Shangqiu 476000, China; 18338709751@163.com
[3] East China Sea Forecasting and Disaster Reduction Center, Ministry of Natural Resources, Shanghai 200136, China; wangleidett727@163.com (L.W.); wuping@ecs.mnr.gov.cn (P.W.)
[4] East China Sea Ecology Center, Ministry of Natural Resources, Shanghai 200136, China; qinyutaox@163.com
* Correspondence: zhangfeng@ecs.mnr.gov.cn

Abstract: In a marine environment, the concentration of chlorophyll is an important indicator of quality, which is also considered an indicator used to predict the marine ecological environment, which is further considered an important means of predicting red tide disasters. Although existing methods for predicting chlorophyll concentration have achieved encouraging performance, there are still two limitations: (i) they primarily focus on the correlation between variables while ignoring negative noise from non-predictive variables and (ii) they are unable to distinguish the impact of chlorophyll from that of non-predictive variables on chlorophyll concentration at future time points. In order to overcome these obstacles, we propose a Multi-Attention Collaborative Network (MACN)-based triangle-structured prediction system. In particular, the MACN consists of two branch networks, with one named **NP-net**, focusing on non-predictive variables, and the other named **T-net**, applied to the target variable. NP-net incorporates variable-distillation attention to eliminate the negative effects of irrelevant variables, and its outputs are used as auxiliary information for T-net. T-net works on the target variable, and both its encoder and decoder are related to NP-net to use the output of NP-net for assistance in learning and prediction. Two actual datasets are used in the experiments, which show that the MACN performs better than various kinds of state-of-the-art techniques.

Keywords: chlorophyll concentration forecasting; multi-attention collaborative; deep neural network; long-term forecasting

check for updates

Academic Editor: Azizur Rahman

Received: 18 December 2024
Revised: 10 January 2025
Accepted: 14 January 2025
Published: 16 January 2025

Citation: Jin, Y.; Zhang, F.; Wang, X.; Wang, L.; Chen, K.; Chen, L.; Qin, Y.; Wu, P. Multi-Step Forecasting of Chlorophyll Concentration with Multi-Attention Collaborative Network. *J. Mar. Sci. Eng.* **2025**, *13*, 151. https://doi.org/10.3390/jmse13010151

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Coastal waters are important areas for high-quality economic development and serve as key pillars for socioeconomic growth. In recent years, high-intensity human activities have had a significant impact on the marine ecological environment, resulting in unprecedented ecological security risks. Chlorophyll concentration, a crucial indicator of primary productivity in oceans, serves as both a measure of eutrophication levels in marine environments and an indicator of overall dynamic changes in the physicochemical properties of seawater. Consequently, the accurate prediction of chlorophyll concentration holds significant value for the early warning of ecological issues such as red tides and green tides, as well as for effective fishery management [1–3].

Currently, there are two primary methods for predicting chlorophyll concentration: those relying on physical and chemical analyses and those utilizing data-driven approaches.

Among these, the traditional method, which relies on physical and chemical analyses, is grounded in the chemical properties of the water body itself. It utilizes complex water dynamics models and physical and chemical prediction models of water quality to analyze and predict chlorophyll concentration [4]. For instance, Jin et al. [5] employed empirical models to predict chlorophyll changes in the Jinhe Reservoir in South Korea. However, due to the diversity of water quality variables in the ocean, establishing an accurate dynamics model is challenging. Furthermore, the incomplete physical parameterization scheme of the model also hampers its prediction accuracy [6,7]. In recent years, with the advancement of deep learning, data-driven methods have become widely used in the field of chlorophyll concentration prediction [8]. To predict chlorophyll concentration in the Yellow Sea and Bohai Sea of China, Yao et al. [9] analyzed and compared various deep learning models.

Despite the capability of deep learning-based prediction methods to extract valuable information from spatiotemporal data and reveal development trends and changing patterns, there remain challenges in accurately predicting chlorophyll concentrations in nearshore areas. **Firstly**, most contemporary methods predominantly concentrate on the interplay among variables, neglecting the discordance of information existing between target variables and those that are non-predictive. In practical situations, incorporating certain non-predictive variables can impede our ability to make accurate judgments. For example, Li et al. [10] discovered that including specific variables in a cascade not only fails to enhance the prediction of future stock prices but actually undermines the accuracy of the predictions. **Secondly**, while they tap into potential patterns over time from a distinct perspective, they overlook the varying temporal significance of target and non-predictive series in the context of prediction. For example, air quality typically improves after rainfall, yet a sudden spike in PM2.5 levels could be attributed to a strong wind that occurred a few hours prior [11]. However, RNNs blindly combine information from the target and non-predictive variables into a hidden state for prediction.

To tackle these challenges, we introduce a novel encoder–decoder architecture known as the **M**ulti-Attention **C**ollaborative **N**etwork (MACN), specifically designed for the multi-step forecasting of chlorophyll concentration. The MACN comprises two branches: NP-net for non-predictive variables and T-net for chlorophyll concentration. T-net employs a modified version of the Long Short-term Memory Network (LSTM), specifically the knowledge-enhanced LSTM (KeLSTM), as both its encoder and decoder. This setup incorporates interactive attention mechanisms and a fusion gate. NP-net, a hierarchical network that integrates our proposed variable-distillation attention mechanism with LSTM, supports T-net by bridging its encoder and decoder. Within T-net, the KeLSTM effectively resolves information conflicts and evaluates temporal influences during both the encoding and decoding processes. We draw the following conclusions about the main contributions of this work:

1. We introduce a novel framework for forecasting chlorophyll concentration, named the MACN. The MACN model increases the forecasting accuracy by closing the information gap that exists between non-predictive and chlorophyll variables. Furthermore, the MACN offers multi-level interpretability for prediction tasks.
2. To minimize the effect of irrelevant variables, we created a hierarchical NP-net with variable-distillation attention. Furthermore, we designed KeLSTM to tackle the information conflict that arises during the computation of chlorophyll and non-predictive variables.
3. Our experimental results on two real datasets demonstrate that the MACN surpasses various baseline models in terms of forecasting performance.

2. Related Work

In this section, we provide a concise overview of some related research in the areas of chlorophyll concentration forecasting and time series forecasting.

Chlorophyll concentration forecasting: Chlorophyll concentration has emerged as a key indicator for predicting red tides, prompting numerous research efforts. Nazeer et al. [12] employed environmental hydrodynamic systems to predict phycocyanin concentration. Park et al. [13] conducted an analysis of coastal waters in Hong Kong using an ecological dynamics model to investigate the relationship between phytoplankton cells and water quality parameters. Additionally, artificial neural networks (ANNs), such as backpropagation networks (BP) [14,15] and LSTM [16], have become widely used for chlorophyll prediction. Yussof et al. [17] employed LSTM and Convolutional Neural Network (CNN) methodologies to forecast chlorophyll concentration levels along the coastline of Sabah, utilizing these predictions to identify and assess red tide occurrences. Mu et al. [18] innovatively applied graph neural networks (GNNs) to quantify causal interdependencies among a diverse array of marine biological variables and seamlessly integrated attention mechanisms within their chlorophyll concentration prediction models. Sun et al. [19] integrated Transformer networks with Fourier analysis to develop the ChloroFormer prediction model, which exhibited remarkable proficiency in capturing both short-term and medium-term dependency patterns within chlorophyll concentration data.

Time series forecasting has seen a recent increase in the use of deep neural networks, especially Deep Belief Networks (DBNs), LSTM networks, and attention-based recurrent neural networks (RNNs) [20,21]. Peng et al. [22] integrated attention mechanisms into their encoder to capture correlations among non-predictive variables. In practical situations, the multi-step forecasting of time series is of great importance as it allows for more informed decision-making concerning future events across multiple time points. Lai et al. [23] proposed a framework that combines long- and short-term time series networks for long-term forecasting by extracting local dependency patterns among variables and identifying long-term trends. Du et al. [24] developed an attention-based encoder–decoder network for multivariate time series multi-step-ahead forecasting, showcasing its superiority over traditional methods such as the Autoregressive Integrated Moving-Average Model (ARIMA), Support Vector Regression (SVR), and Gated Recurrent Units (GRUs). The attention mechanism is a more flexible information soft-selection strategy that can be specially designed according to task requirements. Liu et al. [25] introduced an enhanced attention-based RNN prediction model known as dual-stage, two-phase, attention-based recurrent neural networks, which take into account both temporal dependencies within the series and spatial correlations among variables.

3. Problem Definition and Notations

The objective of chlorophyll multi-step forecasting is to predict future values over multiple time steps based on observed data, which include chlorophyll concentration as well as variables that are non-predictive. Here, we use the symbols $\{\boldsymbol{x}_t\}_{t=1}^{T} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$ and $\{y_t\}_{t=1}^{T} = \{y_1, y_2, \ldots, y_T\}$ to represent non-predictive variables and chlorophyll series, respectively, for the past T time slots. The symbol $\boldsymbol{x}_t = \{x_t^1, x_t^2, \ldots, x_t^n\} \in \mathbb{R}^n (1 \leq t \leq T)$ denotes a vector containing n non-predictive variables at time step t. Simultaneously, we use $y_t \in \mathbb{R}$ to represent the corresponding chlorophyll value at time t. The prediction model offers an estimation of the chlorophyll value for Δ time steps subsequent to T, referred to as $\{\hat{y}_t\}_{t=T+1}^{T+\Delta} = \{\hat{y}_{T+1}, \hat{y}_{T+2}, \ldots, \hat{y}_{T+\Delta}\}$.

$$\{\hat{y}_t\}_{t=T+1}^{T+\Delta} = F(\{\boldsymbol{x}_t\}_{t=1}^{T}, \{y_t\}_{t=1}^{T}) \tag{1}$$

where $F(\cdot)$ is a non-linear mapping function.

4. Proposed MACN Model

Initially, we provide an overview of the proposed MACN framework in this section. Subsequently, we will expand on the technical specifics of each component of the MACN.

4.1. An Overview of the MACN

As depicted in Figure 1, the MACN framework has two branch networks from its top to bottom: T-net and NP-net. T-net is a structured encoder–decoder neural architecture that employs attention mechanisms and operates on the target variable series $\{y_t\}_{t=1}^{T} = \{y_1, y_2, \ldots, y_T\}$. Meanwhile, NP-net functions as an auxiliary network to T-net, focusing on the non-predictive variables $\{\boldsymbol{x}_t\}_{t=1}^{T} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_T\}$. Intuitively, the MACN is a triangular-structured deep learning model that comprises T-net networks, which reinforce chlorophyll semantics from top to bottom, and NP-net networks, which acquire external knowledge. In the MACN, NP-net serves as a secondary network and is connected to both the encoder and decoder of T-net. T-net functions as the backbone network of the MACN model, where KeLSTM converts the output of NP-net into its own auxiliary information. During the encoding phase, KeLSTM extracts valuable features from non-predictive variables and utilizes them to enrich the semantic representation of chlorophyll. In the decoding phase, KeLSTM aids the model in making decisions based on the external knowledge provided by NP-net. Notably, information attention is structurally independent of temporal attention (TAN), which assists T-net in determining the contribution of non-predictive sequences to predictions and capturing the temporal correlation of chlorophyll sequences with the tasks at hand.

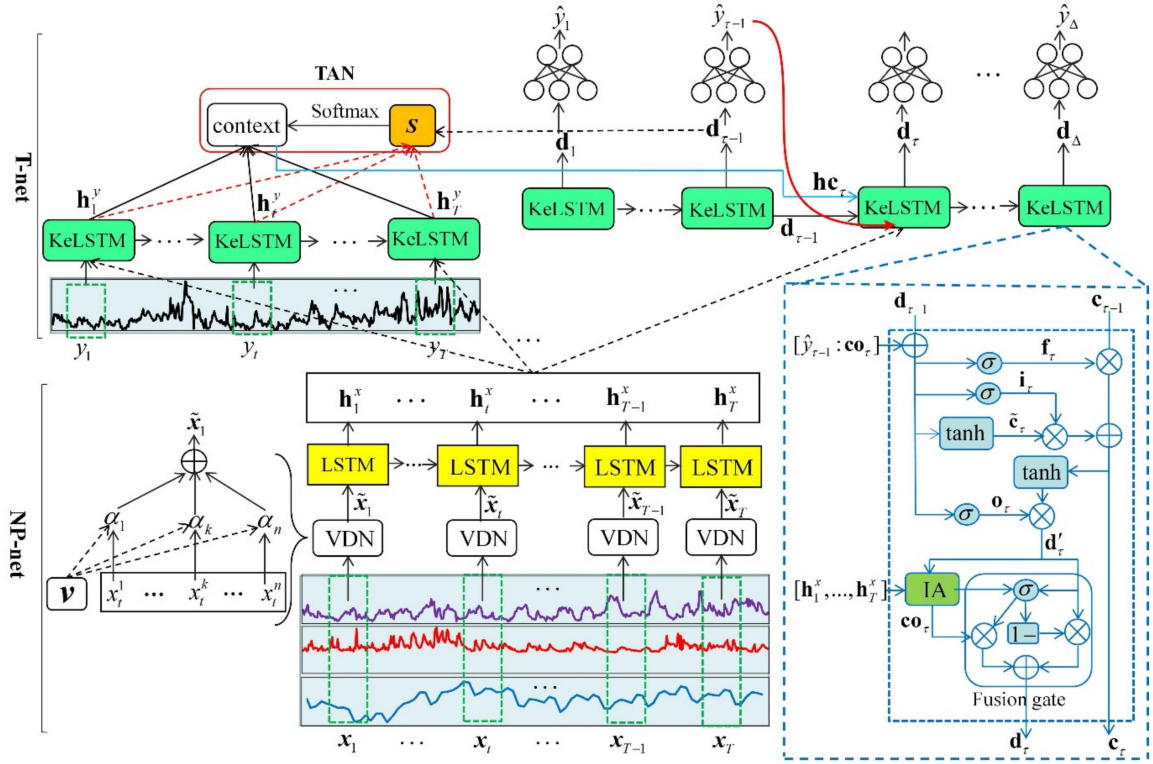


Figure 1. Illustration of the architecture of the MACN.

4.2. NP-Net on Non-Predictive Variables

Variable-distillation attention network (VD): Despite the promising performance of existing attention-based RNNs, they overlook the adverse effects of irrelevant variables. Therefore, we have developed a VD to pinpoint crucial non-predictive variables. For a

given non-predictive series $\boldsymbol{x}_t$, the VD assigns an importance score using a scoring function that involves x_t^k ($1 \leq k \leq n$). The score quantifies the significance of x_t^k relative to $\boldsymbol{x}_t$. In the VD, each variable x_t^k is transformed into an m-dimensional space to derive its hidden representation $\boldsymbol{x}_t$. Following this, these hidden representations are combined to generate the spatial semantics $\widetilde{\boldsymbol{x}}_t$ associated with $\boldsymbol{x}_t$. The specific aggregation process is detailed below:

$$\begin{aligned} \boldsymbol{x}_t^k &= \tanh(\boldsymbol{w}x_t^k + \mathbf{b}_t^k) \\ \alpha_t^k &= \frac{\exp(v\boldsymbol{x}_t^k)}{\sum_{m=1}^{n} \exp(v\boldsymbol{x}_t^m)} \\ \widetilde{\boldsymbol{x}}_t &= \sum_{k=1}^{n} \alpha_t^k \boldsymbol{x}_t^k \end{aligned} \tag{2}$$

where the vector $\boldsymbol{v}$ serves as the variable-level context vector, representing the high-level abstraction of a constant query: 'What constitutes a significant factor for the task?'. Please note that the vector $\boldsymbol{v}$ is initialized at random and its values are learned through the training process. Next, we applied the softmax function to convert the importance score, $v\boldsymbol{x}_t^k$, of variable x_t^k into an attention weight, α_t^k. Ultimately, we consolidated the series of $\{\boldsymbol{x}_t^1, \boldsymbol{x}_t^2, \ldots, \boldsymbol{x}_t^n\}$ into spatial semantics, $\widetilde{\boldsymbol{x}}_t$.

Temporal dependence modeling: LSTM is a versatile and valuable model for learning sequential data. Consequently, we utilized LSTM to capture the temporal dependencies within the series of spatial semantics $\{\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2, \ldots, \widetilde{\boldsymbol{x}}_T\}$.

$$\begin{aligned} \mathbf{i}_t &= \sigma(\mathbf{W}_i\widetilde{\boldsymbol{x}}_t + \mathbf{U}_i\mathbf{h}_{t-1}^{\boldsymbol{x}} + \mathbf{b}_i) \\ \mathbf{f}_t &= \sigma(\mathbf{W}_f\widetilde{\boldsymbol{x}}_t + \mathbf{U}_f\mathbf{h}_{t-1}^{\boldsymbol{x}} + \mathbf{b}_f) \\ \mathbf{o}_t &= \sigma(\mathbf{W}_o\widetilde{\boldsymbol{x}}_t + \mathbf{U}_o\mathbf{h}_{t-1}^{\boldsymbol{x}} + \mathbf{b}_o) \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c\widetilde{\boldsymbol{x}}_t + \mathbf{U}_c\mathbf{h}_{t-1}^{\boldsymbol{x}} + \mathbf{b}_c) \\ \mathbf{h}_t^{\boldsymbol{x}} &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \end{aligned} \tag{3}$$

where the symbol $\mathbf{h}_t^{\boldsymbol{x}} \in \mathbb{R}^m$ represents the hidden state of the LSTM network at time step t, and it is a vector with m dimensions. $\mathbf{i}_t$, $\mathbf{o}_t$, and $\mathbf{f}_t$ correspond to the input gate, output gate, and forget gate of the LSTM network, respectively. $\odot$ represents element-wise multiplication, while σ signifies the logistic function.

4.3. T-Net on Target Variable

While NP-net is effective in mitigating the adverse effects of non-predictive variables, it fails to account for the target variable. To address this, we designed an attention-based encoder–decoder network called T-net, which focuses on the target variable series. Specifically, we meticulously crafted a knowledge-enhanced LSTM (KeLSTM) network that incorporates an interactive attention mechanism and introduces a fusion gate unit. The KeLSTM serves as both the encoder and decoder for T-net. Additionally, we introduced a temporal attention network (TAN) to serve as a bridge between the encoder and decoder of T-net, facilitating the selection of the most relevant items from the encoder for prediction.

Temporal attention network (TAN): The TAN employs a soft-selection strategy, enabling the selection of encoder output items based on their relevance and importance to the decoder. Specifically, when given a source input sequence denoted as $\{z_i\}_{i=1}^n = \{z_1, z_2, \ldots, z_n\}$ and a query vector, $\boldsymbol{q}$, that is pertinent to the task, the temporal attention mechanism computes an importance score for each input item z_i ($1 \leq i \leq n$) using a scoring function. This process can be mathematically represented by the following equation:

$$\begin{aligned} e_i &= \boldsymbol{v}^T \tanh(\mathbf{W}\boldsymbol{z}_i + \mathbf{U}\boldsymbol{q}) \\ \beta_i &= \frac{\exp(e_i)}{\sum_{k=1}^{n} \exp(e_k)} \end{aligned} \tag{4}$$

where $\boldsymbol{v} \in \mathbb{R}^m$,$\mathbf{W} \in \mathbb{R}^{m\times m}$, and $\mathbf{U} \in \mathbb{R}^{m\times m}$ represent learnable parameters. The symbol e_i denotes the importance score assigned to input item $\boldsymbol{z}_i$, while β_i signifies the attention weight obtained after applying the logistic function. The TAN's output is the weighted sum of all items in set $\{\boldsymbol{z}_i\}_{i=1}^n$, with the weights being the attention weights $\{\beta_i\}_{i=1}^n$.

$$\mathbf{c} = \sum_{i=1}^{n} \beta_i \odot \boldsymbol{z}_i \tag{5}$$

where $\mathbf{c} \in \mathbb{R}^m$ is a context vector.

KeLSTM: KeLSTM is a new variant of LSTM, and its structure is shown in Figure 1. Intuitively, KeLSTM incorporated interactive attention (i.e., IA module) and a fusion gate. Consequently, we transformed the output of a standard LSTM network into a candidate state. Here, we employ symbols $\mathbf{h}'_t$ and $\mathbf{h}_t$ to denote the candidate state and the hidden state of KeLSTM at time t, respectively. Figure 1 illustrates that there are two parts to the input of the KeLSTM network at time t: the auxiliary information matrix and the variable value. For simplicity and without losing generality, we will refer to them as χ_t and $\{\boldsymbol{ex}_t\}_{t=1}^n$ = $\{\boldsymbol{ex}_1, \boldsymbol{ex}_2, \ldots, \boldsymbol{ex}_n\}$, respectively. Initially, the KeLSTM network utilizes the input gate $\mathbf{i}_t$, forget gate $\mathbf{f}_t$, and output gate $\mathbf{o}_t$ to produce the candidate state.

$$\begin{aligned}
\mathbf{i}_t &= \sigma(\mathbf{W}_i\chi_t + \mathbf{U}_i\mathbf{h}_{t-1} + \mathbf{b}_i)\\
\mathbf{f}_t &= \sigma(\mathbf{W}_f\chi_t + \mathbf{U}_f\mathbf{h}_{t-1} + \mathbf{b}_f)\\
\mathbf{o}_t &= \sigma(\mathbf{W}_o\chi_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o)\\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c\chi_t + \mathbf{U}_c\mathbf{h}_{t-1} + \mathbf{b}_c)\\
\mathbf{h}'_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t)
\end{aligned} \tag{6}$$

Subsequently, KeLSTM applies an IA module to integrate the auxiliary information matrix $\{\boldsymbol{ex}_t\}_{t=1}^T$ into a single auxiliary context vector. To achieve this, it employs the candidate state $\mathbf{h}'_t$ as the query vector, assessing the significance of each auxiliary vector $\boldsymbol{ex}_t(1 \le t \le T)$.

$$\begin{aligned}
s_i &= \boldsymbol{ex}_i\mathbf{W}(\mathbf{h}'_t)^{\mathrm{T}}\\
\gamma_i &= \frac{\exp(s_i)}{\sum_{k=1}^{n}(s_k)}\\
\boldsymbol{ac} &= \sum_{k=1}^{n} \gamma_i \odot \boldsymbol{ex}_i
\end{aligned} \tag{7}$$

where $(*)^{\mathrm{T}}$ stands for the transpose of a matrix, $\mathbf{W} \in \mathbb{R}^{m\times m}$ serves as a learnable parameter, and $\boldsymbol{ac} \in \mathbb{R}^m$ is designated as the auxiliary context vector. The notations s_i and γ_i are used to indicate the importance score and attention weight, respectively. Ultimately, KeLSTM utilizes a fusion gate unit to produce the hidden state.

$$\begin{aligned}
\mathbf{g}_t &= \sigma(\mathbf{W}_a\boldsymbol{ac} + \mathbf{W}_h\mathbf{h}'_t)\\
\mathbf{h}_t &= \mathbf{g}_t \odot \boldsymbol{ac} + (1 - \mathbf{g}_t) \odot \mathbf{h}'_t
\end{aligned} \tag{8}$$

The logistic sigmoid function maps the value of $\mathbf{g}_t$ to the interval [0,1]. Consequently, when $\mathbf{g}_t$ = 0, the fusion gate effectively disregards the information contained in $\boldsymbol{ac}$.

Encoding process: The MACN model incorporates KeLSTM to develop the encoder for T-net, which is then linked to NP-net. During this stage, the outputs from NP-Net function as supplementary data for the encoder, aiming to bolster the representation of the target variable. Specifically, at time t, the KeLSTM network leverages its three gate units to convert the present target variable and the preceding hidden state into candidate states.

Following this, the KeLSTM network augments these candidate states by incorporating the auxiliary information matrix.

$$\mathbf{h}_t^y = KeLSTM(y_t, \mathbf{h}_{t-1}^y, \{\mathbf{h}_t^x\}_{t=1}^T) \tag{9}$$

where $KeLSTM(*)$ represents KeLSTM, which can be computed according to Equations (6)–(8) with χ_t and $\boldsymbol{ex}_i$ replaced by the newly derived y_t and $\mathbf{h}_t^x$. This symbol denotes the hidden state of the KeLSTM network at time t. Consequently, we obtain a hidden representation $\left\{\mathbf{h}_t^y\right\}_{t=1}^T$ that is associated with the target variable series $\{y_t\}_{t=1}^T$.

Decoding process: For descriptive clarity, we use the symbol $\mathbf{d}_\tau \in \mathbb{R}^m (1 \leq \tau \leq \Delta)$ to represent the hidden state of the decoder unit τ. Given the hidden state series $\left\{\mathbf{h}_t^y\right\}_{t=1}^T$ and the previous decoder state $\mathbf{d}_{\tau-1}$, the temporal attention module combines the sequence into a vector.

$$\mathbf{hc}_\tau = TAN(\mathbf{h}_t^y, \mathbf{d}_{\tau-1}) \tag{10}$$

where $TAN(*)$ denotes temporal attention, which can be computed according to Equation (4), with $\boldsymbol{z}_i$ and $\boldsymbol{q}$ replaced by the newly derived $\mathbf{h}_t^y$ and $\mathbf{d}_{\tau-1}$. Subsequently, the KeLSTM network obtains the hidden state $\mathbf{d}_\tau$ by integrating the context vector $\mathbf{hc}_\tau$ with the previous hidden state $\mathbf{d}_{\tau-1}$. Specifically,

$$\mathbf{d}_\tau = KeLSTM([\hat{y}_{\tau-1} : \mathbf{hc}_\tau], \mathbf{d}_{\tau-1}, \{\mathbf{h}_t^x\}_{t=1}^T) \tag{11}$$

where $[:]$ represents the concatenation operation, and $\hat{y}_{\tau-1}$ is the predicted value of the decoder unit $\tau - 1$. The entire process can be computed according to Equations (6)–(8).

Task learning: We utilize a multilayer perceptron (MLP) as the task learning layer of our model to determine the predicted outcomes of the decoder unit τ. The detailed calculation process is outlined below:

$$\hat{y}_\tau = \mathbf{W}_p \mathbf{d}_\tau \tag{12}$$

where $\mathbf{W}_p \in \mathbb{R}^m$ is a learnable parameter.

5. Experimental Results and Analyses

5.1. Datasets and Baseline Approaches

In this study, we aimed to predict chlorophyll concentration using monitoring data gathered from the coastal waters of Xiamen, Fujian Province, China. The Xiamen Sea area encompasses Tongan Bay (Area 2), the Western Waters (Area 3), and the Eastern Waters (Area 4), all surrounding Xiamen Island (Area 1). Tongan Bay and the Western Waters are delineated by the Gaoji seawall and interconnected via a red line, as illustrated in Figure 2. For our experiments, we selected two buoy monitoring datasets provided by the Fujian Ocean Forecasting Station. These datasets consist of continuous measurements of meteorological conditions, water quality, and nutritional status variables, collected by various buoys equipped with diverse sensors at a 30 min time resolution.

WWA-Data: This dataset, collected by the Marine Ecological Environment Monitoring Station in the Western Waters from January 2009 to August 2011, comprises 7107 time series. It encompasses nine variables: chlorophyll (Chl), sea surface temperature (SST), dissolved oxygen (DO), saturated dissolved oxygen (SDO), tide, air temperature (Air_temp), standard atmospheric pressure (SAP), and two meteorological wind components labeled as Wind_u and Wind_v.

TAW-Data: This dataset originates from Tongan Bay and covers a monitoring period ranging from January 2009 to July 2017, during which 8733 time series were documented.

Each of these sequences contains, in addition to the variables mentioned in the WWA-Data, turbidity (Turb) and PH.

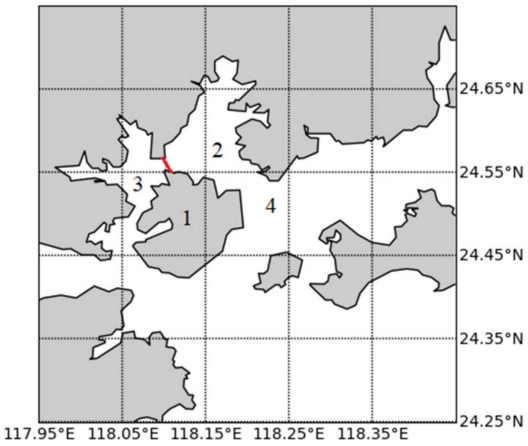


Figure 2. Offshore waters of Xiamen City. The red line represents Gaoji Dam.

During the experiment, chlorophyll concentration served as the target variable. The model was trained using the first 90% of the dataset, while its performance was evaluated using the remaining 10% of the data. The evaluation process comprised two parts. In the first part, the MACN model was benchmarked against currently advanced baseline methods to investigate its unique characteristics. The second part involved an ablation experiment, where the MACN model was compared with its degraded version to assess the role and function of its individual components.

MTSMFF: the multivariate time series forecasting framework (MTSMFF) introduced by Du et al. [24] has demonstrated its superiority in comparison to conventional methodologies, including ARIMA, SVR, LSTM, and GRUs.

DA-RNN: Peng [22] incorporated attention mechanisms into both the encoder and decoder. In our study, we tailored a dual-stage attention-based recurrent neural network (DA-RNN) for long-term predictions by adopting a direct strategy.

DSTP-RNN: the dual-stage, two-phase, attention-based recurrent neural network (DSTP-RNN), as proposed by Liu et al. [25], exhibits excellent performance in the long-term prediction of multivariate time series.

TPA-LSTM: the temporal pattern attention LSTM (TPA-LSTM) method, specifically designed for multivariate time series [26], utilizes a series of filters to transform the time series into different 'frequency domains' and extract time-invariant temporal patterns.

DyAt-Nets: Muralidhar et al. [27] designed dynamic attention networks (DyAt-Nets), which incorporate previous decoder states into the current decoder unit.

DA-TLSTM: Hu et al. [28] proposed dual-stage, attention-based T-LSTM (DA-TLSTM) as a long-term forecasting model for multivariate time series, capable of extracting the influence of temporal correlation information. In our study, we implemented DA-TLSTM for the multi-step forecasting of chlorophyll concentration.

MACN-Dv1: This model employs NP-net and LSTM to build its encoder and decoder, respectively, with TAN serving as the intermediary that connects the two. Structurally, MACN-Dv1 can also be viewed as a variant of the MTSMFF, enhanced by the addition of a VD module to its encoder.

MACN-Dv2: MACN-Dv2 eliminates the VD module from the auxiliary network, NP-net. As a result, MACN-Dv2 can only model the temporal correlation of non-predictive sequences and chlorophyll sequences relevant to prediction tasks independently, without the capability to alleviate the adverse effects of non-predictive variables.

5.2. Parameter Setting and Performance Evaluation

Parameter setting: We set the learning rate for all methods to 0.0001. Consequently, all models except TPA-LSTM had only one hyperparameter to determine, which was the size of the hidden state. To ensure consistency in the context feature dimensions, we used the same hidden state size for all LSTM networks in the model. We performed a grid search for the hidden state size of the LSTM networks over the range {15, 20, 25, 30, 35, 40}. Additionally, for TPA-LSTM, we conducted a grid search for the size of the convolution kernel over the range {24, 48, 64, 128}. We set the size of the time window to 24 (i.e., T = 24). In this study, we compared our model with previous state-of-the-art methods and evaluated all the methods on two datasets with $\Delta \in \{1, 3, 6, 12, 24\}$.

Performance evaluation: The predictive performance of the model was evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The RMSE quantifies the average squared deviation of the predicted results, while the MAE measures the average absolute difference between the predicted and original values. The closer the values of the RMSE and MAE are to 0, the more accurate the model's predictions.

$$\begin{aligned} MAE &= \frac{1}{\Delta}\sum_{t=1}^{\Delta}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{y}_t^i - y_t^i|\right) \\ RMSE &= \frac{1}{\Delta}\sum_{t=1}^{\Delta}\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_t^i - y_t^i)^2}\right) \end{aligned} \tag{13}$$

where Δ is the size of the predictive horizon, and N is the number of samples. y_t^i is the real value and $\hat{y}_t^i$ is the corresponding predicted value.

5.3. Experimental Results and Analysis

5.3.1. Comparison with Baselines

To ensure a fair comparison, the input sequence length for all models was set to 24, and the output sequence length of the model was incrementally increased according to the set {1, 3, 6, 12, 24}. This implies that as the value of Δ increases from 1 to 24, the model progressively predicts chlorophyll concentration for the subsequent 30 min up to 720 minutes. To visually assess the performance of the MACN model against the baseline method, Tables 1 and 2 present the results for each method based on different forecast horizons, with the best performance emphasized in bold.

Tables 1 and 2 illustrate the application of chlorophyll concentration prediction in various sea areas, demonstrating that the MACN model outperforms the baseline method for chlorophyll prediction. Additionally, as the value of Δ increases within the set {1, 3, 6, 12, 24}, both the MACN model and the baseline method show a decline in predictive performance. A time series is a sequence of random variables ordered by the time of their occurrence, typically obtained by sampling a phenomenon at regular intervals. Consequently, time series data represent a highly non-linear and time-varying process, with uncertainty intensifying as the time intervals expand. Therefore, as a time point becomes more distant from the observation period (time window), prediction becomes more challenging, resulting in a reduction in model performance as the forecast horizon increases. Based on the performance data presented in Tables 1 and 2, the MACN model exhibited the best performance in 80% of the application cases (16 out of 20). This indicates that an increase in the aa value has a relatively minor adverse effect on the MACN model's performance.

Furthermore, utilizing the application scenario where $\Delta = 24$ as an example, this work visualizes the aforementioned method to evaluate and compare the performance of the MACN against the baseline method across various prediction points, as illustrated in Figures 3 and 4. In these figures, the x-axis values correspond to different prediction points

within a specified prediction horizon, while the y-axis values represent the Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) achieved by the different methods at those specific prediction points. For example, at a point where the x-axis value is 1, the red curve in Figure 3 indicates the predictive capability of the MACN at the initial position in the model's output sequence, assuming a sequence length of 24. Intuitively, when compared with other deep learning-based baseline methods such as MTSMFF, DA-RNN, DSTP-RNN, TPA-LSTM, DyAt-Nets, and DA-TLSTM, the MACN presented in this paper demonstrates an advantage in the practical application of chlorophyll concentration prediction. Upon analyzing the structure of each model, it is evident that the MACN outperforms the other methods due to the following reasons: (i) In contrast to modeling methods that handle all input variables in a unified manner, the MACN's strategy of modeling chlorophyll and non-predictive sequences separately allows for better capturing the interactions between them and more accurate representations of real future scenarios. (ii) The variable-distillation attention network effectively mitigates the negative effects of non-predictive variables. (iii) KeLSTM in T-net successfully captures the potential interactions between past chlorophyll sequences, non-predicted sequences, and upcoming predictions. In particular, it identifies subtle differences between these two potential interactions, specifically the out-of-sync phenomenon regarding their temporal correlation with the prediction task.

Table 1. Performance of different methods based on WWA_Data in different prediction horizons.

Methods	Horizon = 1		Horizon = 3		Horizon = 6		Horizon = 12		Horizon = 24	
	MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
MTSMFF	0.4285	0.5685	0.5698	0.7732	0.8011	0.9529	0.8885	1.0393	0.9689	1.1518
DA-RNN	0.3670	**0.4922**	0.4779	**0.6340**	0.6164	0.7881	0.6399	0.8716	0.7652	0.9619
DSTP-RNN	0.3864	0.5690	0.4787	0.7014	**0.5330**	0.7228	0.5758	0.8032	0.7517	0.9017
TPA-LSTM	0.4307	0.5613	0.5975	0.7468	0.6656	0.8323	0.7181	0.8925	0.7846	0.9765
DyAt-Nets	0.4285	0.5934	0.5134	0.7109	0.5902	0.7882	0.7680	0.9482	0.8205	0.9669
DA-TLSTM	0.4007	0.5799	0.4812	0.7042	0.5461	0.7444	0.6171	0.7826	0.6563	0.8120
MACN	**0.3655**	0.4928	**0.4749**	0.6965	0.5353	**0.6743**	**0.5513**	**0.7115**	**0.6212**	**0.7725**

Table 2. Performance of different methods based on TAW_Data in different prediction horizons.

Methods	Horizon = 1		Horizon = 3		Horizon = 6		Horizon = 12		Horizon = 24	
	MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
MTSMFF	0.4792	0.8051	0.5800	0.9029	0.7001	1.0567	0.7917	1.1341	0.8412	1.2529
DA-RNN	0.4847	0.8495	0.6639	0.9663	0.7678	1.1013	0.8168	1.1835	0.8403	1.2097
DSTP-RNN	0.4589	**0.8004**	0.5849	0.9062	0.6338	**0.9912**	0.7189	1.0568	0.7515	**1.0740**
TPA-LSTM	0.4918	0.8039	0.6130	0.9038	0.6983	1.0196	0.7689	1.1021	0.8267	1.1716
DyAt-Nets	0.5160	0.9042	0.6212	0.8685	0.7023	1.0020	0.7572	1.1203	0.7915	1.3334
DA-TLSTM	0.4492	0.8035	0.5442	0.8083	**0.6239**	1.0246	0.6911	1.1179	0.7215	1.1708
MACN	**0.4425**	0.8163	**0.5233**	**0.7968**	0.6344	1.0125	**0.6578**	**1.0369**	**0.6760**	1.1127

***t*-test analysis:** The significance test is a widely utilized method in hypothesis testing, primarily used to identify whether a statistically significant difference exists between the experimental and control groups in scientific experiments. To precisely determine if there was a notable difference in prediction errors between the MACN model and the baseline method, this paper conducted a paired two-tailed *t*-test for both methods. The outcomes of this hypothesis test are summarized in Table 3. At a significance level of $\alpha = 0.05$, if the *t*-test value fulfills the condition $p \leq 0.05$, it can be concluded that there is a statistically significant difference in the prediction error between the MACN model and the baseline method. It is important to note that the outcomes of the T-hypothesis test can be influenced by sample size. Therefore, this paper also calculated and compared the average RMSE of the MACN and baseline methods from various prediction perspectives. The findings suggest that, at a statistical significance level of 5%, the MACN proposed in this paper

surpasses nearly all baseline methods. Although the *t*-test results using the WWA_Data dataset suggest no statistically significant difference in predictive performance between DA-TLSTM and the MACN, the MACN exhibits a lower average RMSE value. This further confirms that the MACN outperforms DA-TLSTM. In summary, the statistical analysis based on the *t*-test demonstrates that the MACN offers superior multi-step prediction performance compared to the baseline methods.

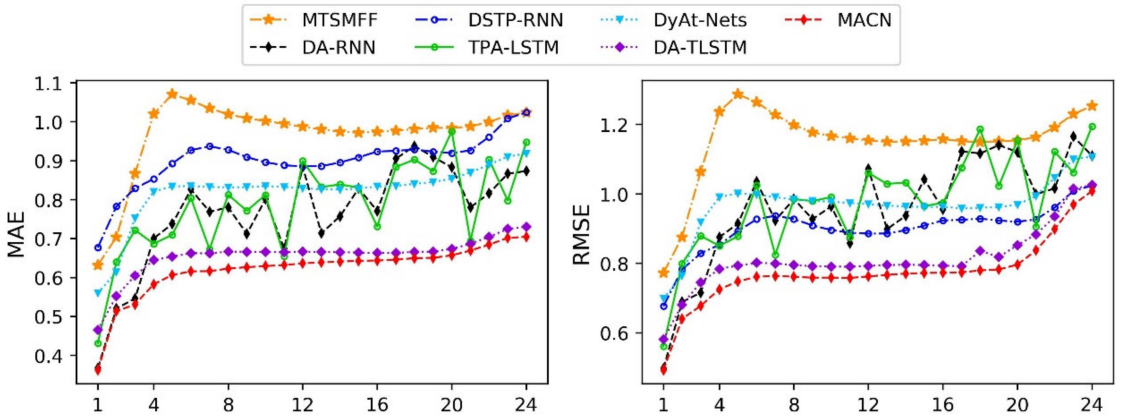


Figure 3. Performance of different methods based on WWA-Data dataset when $\Delta = 24$.

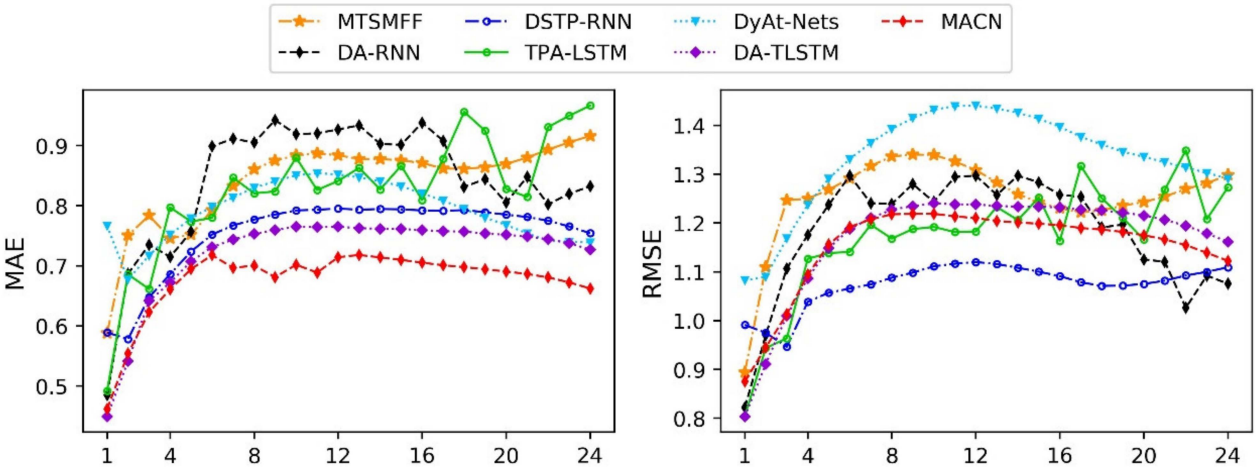


Figure 4. Performance of different methods based on TAW-Data dataset when $\Delta = 24$.

Table 3. *t*-test results for baseline method and MACN, with confidence level for *t*-test scores set at $\alpha = 0.05$.

Methods	WWA_Data			TAW_Data		
	***p*-Value**	**T-Statistic**	**avg.RMSE**	***p*-Value**	**T-Statistic**	**avg.RMSE**
MTSMFF	0.0000	−8.4810	0.8971	0.0000	−5.7623	1.0303
DA-RNN	0.0338	−3.4938	0.7496	0.0204	−2.3238	1.0621
DSTP-RNN	0.0065	−3.7550	0.7396	0.0005	−3.5041	0.9657
TPA-LSTM	0.0231	−2.1301	0.8019	0.0000	−3.9811	1.0002
DyAt-Nets	0.0001	−5.4920	0.8015	0.0010	−3.4610	1.0457
DA-TLSTM	0.0626	−2.0822	0.7246	0.0112	−2.4080	0.9851
MACN	--	--	0.6695	--	--	0.9550

5.3.2. Ablation Study of Model Components

To assess the effectiveness of the VDN module and KeLSTM, this study introduces two simplified versions of the MACN, namely MACN-Dev1 and MACN-Dev2, by sequentially removing one key component from the original model. Section 5.1 provides a comprehensive description of these simplified model structures. Using these degraded versions, this study qualitatively evaluates the effectiveness of each component and the overall structure of the MACN. Ablation experiments were conducted based on chlorophyll datasets from two sea areas, and the results for the MACN model and its degraded versions in various prediction fields are presented in Tables 4 and 5. The best performance in each case is highlighted in bold.

Table 4. Ablation experiments based on TAW-Data dataset.

Methods	Horizon = 1		Horizon = 3		Horizon = 6		Horizon = 12		Horizon = 24	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MTSMFF	0.4792	0.8051	0.5800	0.9029	0.7001	1.0567	0.7917	1.1341	0.8412	1.2529
MACN-Dv1	0.4683	0.8448	0.6003	0.9255	0.6621	**1.0089**	0.7082	1.1022	0.7698	1.1670
MACN-Dv2	0.4722	**0.7922**	0.6095	0.9482	0.6827	1.0525	0.7237	1.0625	0.7555	1.1938
MACN	**0.4425**	0.8163	**0.5233**	**0.7968**	**0.6344**	1.0125	**0.6578**	**1.0069**	**0.6760**	**1.1127**

Table 5. Ablation experiments based on WWA-Data dataset.

Methods	Horizon = 1		Horizon = 3		Horizon = 6		Horizon = 12		Horizon = 24	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MTSMFF	0.4285	0.5685	0.5698	0.7732	0.8011	0.9529	0.8885	1.0393	0.9689	1.1518
MACN-Dv1	0.4062	0.5989	0.5680	0.7068	0.6421	0.7874	0.7180	0.8502	0.7642	0.8809
MACN-Dv2	0.3988	0.5889	0.5782	0.7980	0.6173	0.7708	0.6952	0.8492	0.7483	0.9078
MACN	**0.3655**	**0.5488**	**0.4749**	**0.6965**	**0.5353**	**0.6743**	**0.5513**	**0.7115**	**0.6212**	**0.7746**

The effectiveness of the MACN model structure: To evaluate the effectiveness of distinguishing between the influence of the target sequence and the explanatory sequence on the prediction task in the time dimension and to thereby confirm the validity of the model structure, two sets of comparative experiments were carried out: (i) MACN-Dv1 versus MACN and (ii) MACN-Dv2 versus MTSMFF. The results presented in Tables 4 and 5 show that the MACN outperforms MACN-DV1 across two datasets, while MACN-Dv2 outperforms the MTSMFF, except for specific locations (horizons) within the TAW-Data dataset. These results highlight the benefits of modeling the target sequence and the explanatory sequence independently, reinforcing the effectiveness of the MACN model structure compared to integrating all variables simultaneously into the model.

The effectiveness of the VD module: To evaluate the effectiveness of the VD, two compelling comparative experiments were conducted: (i) MACN-Dv2 versus MACN and (ii) MACN-Dv1 versus MTSMFF. As shown in Tables 4 and 5, the MACN outperformed MACN-Dv2 across both datasets. Furthermore, MACN-Dv2 demonstrated superior performance compared to the MTSMFF. As described in Section 5.1, both MACN-Dv1 and the MACN incorporate VD modules prior to the encoders of the MTSMFF and MACN-Dv2, respectively, yet they achieve better results. Based on these experimental findings, it can be concluded that variable-distillation attention is effective in improving multivariate time series prediction.

In Figure 5, we showcase the visualization of the weight distribution associated with the VD utilizing both the WWA-Data and TAW-Data datasets. The x-axis depicts various time steps, and the y-axis indicates the weights assigned to different variables at those specific time steps. Variables that exhibit higher weight values signify a more

significant influence on the predictions made by the MACN. It is evident that the influence of variables on prediction tasks fluctuates over time. The MACN not only distinguishes between the predictive importance of each variable but also captures dynamic shifts in this importance. As illustrated in Figure 5, for both the WWA-Data and TAW-Data, the VD assigns relatively higher weights to seawater temperature (Temp), air temperature (Air-Temp), and tide (Tide), which is consistent with the findings reported in [29,30]. Consequently, the explanatory information provided by the MACN for prediction tasks is deemed reliable. Furthermore, Figure 5 reveals an additional insight: different combinations of variables result in variations in their contributions to the prediction task. For instance, in the WWA-Data, barometric pressure (Press) plays a significant role in predicting chlorophyll concentration. However, when the PH and Turb variables are incorporated into the TAW-Data, barometric pressure exhibits a weaker correlation with the prediction of chlorophyll concentration. This underscores the MACN's capability to select the variables that are most pertinent to the task based on the inherent characteristics of the data.

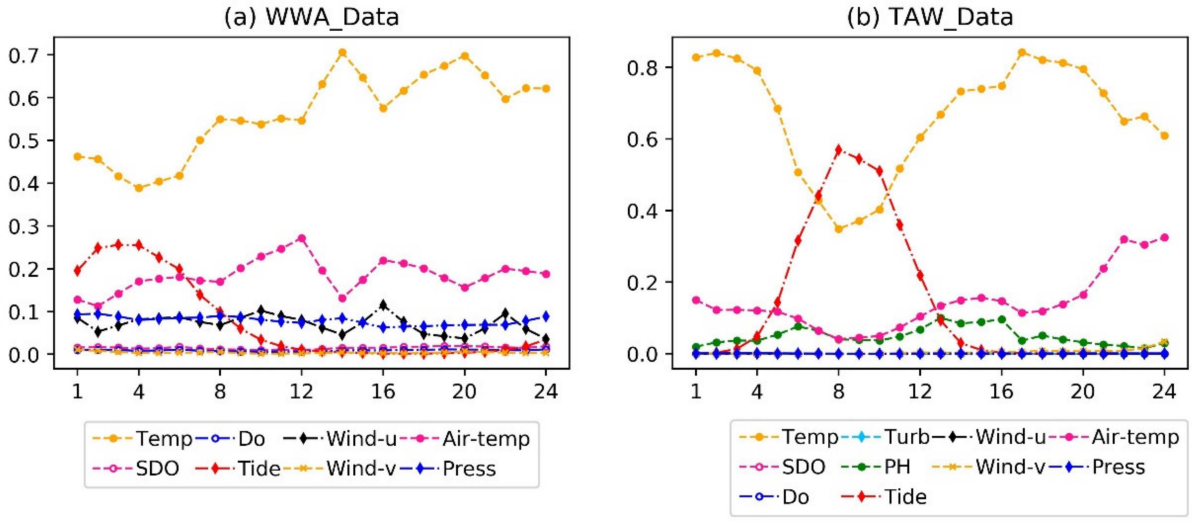


Figure 5. Visualization of variable attention weights for WWA-Data and TAW-Data.

The effectiveness of KeLSTM: To assess the effectiveness of KeLSTM, we compared the performance of the MACN with its downgraded variant, MACN-Dv1. As evident from Tables 4 and 5, the MACN, when paired with KeLSTM, surpassed MACN-Dv1. To further validate the efficacy of KeLSTM, we offer visualizations of the weight distribution of information attention during the prediction process, along with a comparison to the weight distribution of temporal attention, as depicted in Figure 6. In these figures, the coordinates (x_k, y) are used, where x_k represents the k-th observation time step within the observation period (or time window), and y denotes the attention weight at that specific time step. Specifically, the curves labeled 'IA-in-EN' and 'IA-in-DE' illustrate the influence of the historical, non-predicted sequence $\boldsymbol{x}_t (1 \leq t \leq 24)$ on the chlorophyll, y_{24}, and prediction results, y_{25}, respectively. Here, $\boldsymbol{x}_t = (x_t^1, x_t^2, \ldots, x_t^n)$ refers to a vector composed of n non-predictor variables at a given time step t. The curve labeled "TAttention" depicts the contribution of the chlorophyll sequence to the prediction y_{25}.

As shown in Figure 6, the curves labeled 'IA-in-EN' and 'IA-in-DE' exhibit slight variations but follow similar trends, confirming the effectiveness and stability of information attention. Additionally, the visualization results indicate that the volatility of the "TAttention" curve differs from that of the "IA in DE" curve. This observation implies that T-net effectively captured the nutritional distinctions within the temporal correlations. However, considering the predictive performance summarized in Tables 3 and 4,

these experiments underestimate the significance of independently modeling targets and non-predictive series.

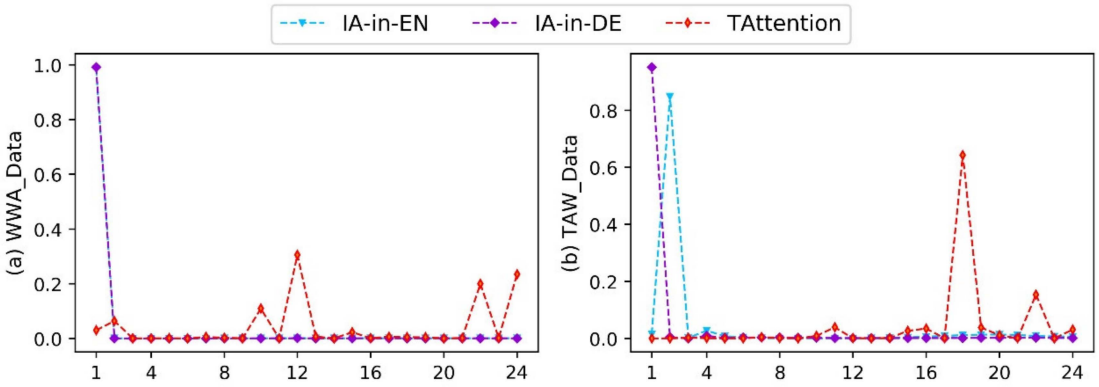


Figure 6. Weight distribution of information attention and temporal attention when $\Delta = 1$.

This paper provides a visual representation of the output produced by the gated fusion unit incorporated within the KeLSTM framework in Figure 7. During implementation, the LSTM unit is set to have a size of 15, resulting in a 15-dimensional vector as the output of the gating unit. The coordinates indicated (x_n,y_k) refer to a specific time step, while x_n represents the value corresponding to a particular dimension within the output vector of the gated fusion unit. As depicted in Figure 7, only a few entries in the visualization legend are close to 0. This suggests that the gated fusion unit does not fully discard the information from the explanatory sequence but instead selectively integrates it into the candidate hidden states of the KeLSTM network. Consequently, this finding implies that the fusion gate preserves, to a certain extent, the informational content of the explanatory sequence and utilizes this information to enhance the potential representation of the target sequence.

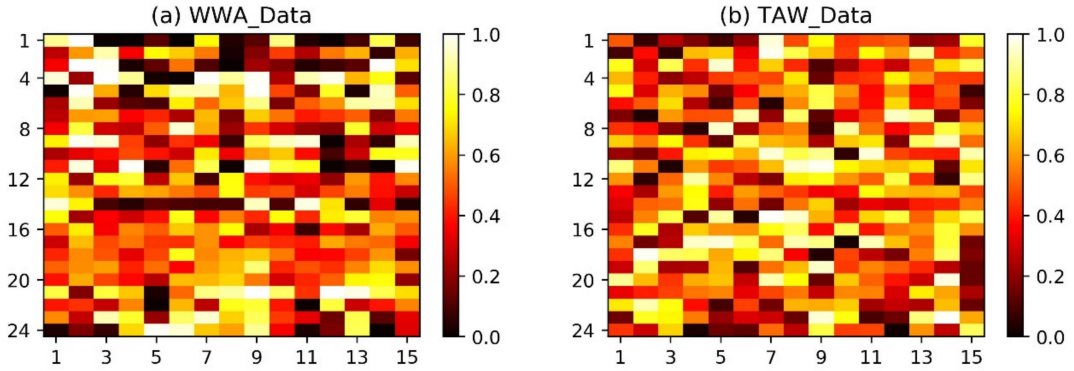


Figure 7. Visualization of parameter distribution of fusion gate based on WWA-Data and TAW-Data.

6. Conclusions

Chlorophyll forecasting remains a significant yet unresolved challenge due to the intricacies of marine environments. In this study, we introduced the MACN, an innovative multivariate time series forecasting model that comprises two branch networks: T-net and NP-net. The MACN is structured as a triangular encoder–decoder network, designed to mitigate information conflicts between chlorophyll and non-predictive variables and to differentiate between their temporal impacts on predictions. The visualization of attention weights demonstrates the model's excellent interpretability. Furthermore, our findings reveal that the effects of the chlorophyll series and non-predictive series on the tasks are not synchronized. Despite its impressive performance and interpretability in chlorophyll prediction, the MACN model fails to account for the mutation phenomenon associated with

chlorophyll prediction and demands substantial computational resources. In our future endeavors, we aim to delve deeper into leveraging mutation information within time series data to enhance prediction accuracy while effectively minimizing resource consumption.

Author Contributions: Methodology, L.W.; Software, L.C. and Y.Q.; Validation, Y.J., K.C., L.C. and Y.Q.; Formal analysis, F.Z.; Investigation, X.W.; Data curation, Y.J., Y.Q. and P.W.; Writing—original draft, F.Z.; Writing—review & editing, X.W., L.W., L.C. and P.W.; Visualization, X.W. and K.C.; Supervision, K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Program on Key Research Project of China (2024YFC3109004).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to reasons related to data ownership.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, K.; Zhao, X.; Xue, J.; Mo, D.; Zhang, D.; Xiao, Z.; Yang, W.; Wu, Y.; Chen, Y. The temporal and spatial variation of chlorophyll a concentration in the China Seas and its impact on marine fisheries. *Front. Mar. Sci.* **2023**, *10*, 1212992. [CrossRef]
2. Liu, N.; Chen, S.; Cheng, Z.; Wang, X.; Xiao, Y.; Xiao, L.; Gong, Y.; Wang, T.; Zhang, X.; Liu, S. Long-term prediction of sea surface chlorophyll-a concentration based on the combination of spatio-temporal features. *Water Res.* **2022**, *211*, 118040.
3. Li, H.; Li, X.; Song, D.; Nie, J.; Liang, S. Prediction on daily spatial distribution of chlorophyll-a in coastal seas using a synthetic method of remote sensing, machine learning and numerical modeling. *Sci. Total Environ.* **2024**, *910*, 168642. [CrossRef]
4. Ham, Y.G.; Joo, Y.S.; Park, J.Y. Mechanism of skillful seasonal surface chlorophyll prediction over the southern Pacific using a global earth system model. *Clim. Dyn.* **2021**, *56*, 45–64. [CrossRef]
5. Jin, S.-H.; Jargal, N.; Khaing, T.T.; Cho, M.J.; Choi, H.; Ariunbold, B.; Donat, M.G.; Yoo, H.; Mamun, M.; An, K.-G. Long-term prediction of algal chlorophyll based on empirical models and the machine learning approach in relation to trophic variation in Juam Reservoir, Korea. *Heliyon* **2024**, *10*, e31643. [CrossRef] [PubMed]
6. Ying, C.; Xiao, L.; Xueliang, Z.; Wenyang, S.; Chongxuan, X. Marine chlorophyll-a prediction based on deep auto-encoded temporal convolutional network model. *Ocean Model.* **2023**, *186*, 102263. [CrossRef]
7. Wu, Q.; Wang, X.; He, Y.; Zheng, J. The Relationship between Chlorophyll Concentration and ENSO Events and Possible Mechanisms off the Changjiang River Estuary. *Remote Sens.* **2023**, *15*, 2384. [CrossRef]
8. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [CrossRef]
9. Yao, L.; Wang, X.; Zhang, J.; Yu, X.; Zhang, S.; Li, Q. Prediction of Sea Surface Chlorophyll-a Concentrations Based on Deep Learning and Time-Series Remote Sensing Data. *Remote Sens.* **2023**, *15*, 4486. [CrossRef]
10. Li, H.; Shen, Y.; Zhu, Y. Stock price prediction using attention-based multi-input LSTM. In Proceedings of the Asian Conference on Machine Learning, PMLR, Beijing, China, 14–16 November 2018; pp. 454–469.
11. Yi, X.; Zhang, J.; Wang, Z.; Li, T.; Zheng, Y. Deep distributed fusion network for air quality prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 965–973.
12. Nazeer, M.; Wong, M.S.; Nichol, J.E. A new approach for the estimation of phytoplankton cell counts associated with algal blooms. *Sci. Total Environ.* **2017**, *590*, 125–138. [CrossRef]
13. Park, Y.; Pyo, J.; Kwon, Y.S.; Cha, Y.; Lee, H.; Kang, T.; Cho, K.H. Evaluating physico-chemical influences on cyanobacterial blooms using hyperspectral images in inland water, Korea. *Water Res.* **2017**, *126*, 319–328. [CrossRef] [PubMed]
14. Abbas, A.; Park, M.; Baek, S.S.; Cho, K.H. Deep learning-based algorithms for long-term prediction of chlorophyll-a in catchment streams. *J. Hydrol.* **2023**, *626*, 130240. [CrossRef]
15. Ye, M.; Li, B.; Nie, J.; Wen, Q.; Wei, Z.; Yang, L.-L. Graph Convolutional Network Assisted SST and Chl-a Prediction with Multi-Characteristics Modeling of Spatio-Temporal Evolution. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14.
16. Yu, Z.; Yang, K.; Luo, Y.; Shang, C. Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network. *J. Hydrol.* **2020**, *582*, 124488. [CrossRef]
17. Yussof, F.N.; Maan, N.; Md Reba, M.N. LSTM networks to improve the prediction of harmful algal blooms in the west coast of Sabah. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7650. [CrossRef]

18. Mu, B.; Qin, B.; Yuan, S.; Wang, X.; Chen, Y. PIRT: A physics-informed red tide deep learning forecast model considering causal-inferred predictors selection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
19. Sun, X.; Yan, D.; Wu, S.; Chen, Y.; Qi, J.; Du, Z. Enhanced forecasting of chlorophyll-a concentration in coastal waters through integration of Fourier analysis and Transformer networks. *Water Res.* **2024**, *263*, 122160. [CrossRef]
20. Lim, B.; Zohren, S. Time-series forecasting with deep learning: A survey. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200209. [CrossRef]
21. Zhang, X.; Qian, B.; Li, Y.; Cao, S.; Davidson, I. Context-aware and time-aware attention-based model for disease risk prediction with interpretability. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3551–3562. [CrossRef]
22. Peng, J.; Kimmig, A.; Wang, J.; Liu, X.; Niu, Z.; Ovtcharova, J. Dual-stage attention-based long-short-term memory neural networks for energy demand prediction. *Energy Build.* **2021**, *249*, 111211. [CrossRef]
23. Lai, G.; Chang, W.C.; Yang, Y.; Liu, H. Modeling long-and short-term temporal patterns with deep neural networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 95–104.
24. Du, S.; Li, T.; Yang, Y.; Horng, S.-J. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing* **2020**, *388*, 269–279. [CrossRef]
25. Liu, Y.; Gong, C.; Yang, L.; Chen, Y. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. *Expert Syst. Appl.* **2020**, *143*, 113082. [CrossRef]
26. Shih, S.Y.; Sun, F.K.; Lee, H. Temporal pattern attention for multivariate time series forecasting. *Mach. Learn.* **2019**, *108*, 1421–1441. [CrossRef]
27. Muralidhar, N.; Muthiah, S.; Ramakrishnan, N. Dyat nets: Dynamic attention networks for state forecasting in cyberphysical systems. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 3180–3186. [CrossRef]
28. Hu, J.; Zheng, W. A deep learning model to effectively capture mutation information in multivariate time series prediction. *Knowl.-Based Syst.* **2020**, *203*, 106139. [CrossRef]
29. Liu, X.; Feng, J.; Wang, Y. Chlorophyll a predictability and relative importance of factors governing lake phytoplankton at different timescales. *Sci. Total Environ.* **2019**, *648*, 472–480. [CrossRef]
30. Cha, Y.K.; Cho, K.H.; Lee, H. The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. *Water Res.* **2017**, *124*, 11–19. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.