remote sensing

MDPI

Article

OKG-ConvGRU: A Domain Knowledge-Guided Remote Sensing Prediction Framework for Ocean Elements

Renhao Xiao, Yixiang Chen *, Lizhi Miao, Jie Jiang, Donglin Zhang and Zhou Su

School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; b22100514@njupt.edu.cn (R.X.); miaolz@njupt.edu.cn (L.M.); jiangj@njupt.edu.cn (J.J.); b22100516@njupt.edu.cn (D.Z.); b22100615@njupt.edu.cn (Z.S.)

* Correspondence: chenyixiang@njupt.edu.cn

Abstract

Accurate prediction of key ocean elements (e.g., chlorophyll-a concentration, sea surface temperature, etc.) is imperative for maintaining marine ecological balance, responding to marine disaster pollution, and promoting the sustainable use of marine resources. Existing spatio-temporal prediction models primarily rely on either physical or data-driven approaches. Physical models are constrained by modeling complexity and parameterization errors, while data-driven models lack interpretability and depend on high-quality data. To address these challenges, this study proposes OKG-ConvGRU, a domain knowledge-guided remote sensing prediction framework for ocean elements. This framework integrates knowledge graphs with the ConvGRU network, leveraging prior knowledge from marine science to enhance the prediction performance of ocean elements in remotely sensed images. Firstly, we construct a spatio-temporal knowledge graph for ocean elements (OKG), followed by semantic embedding representation for its spatial and temporal dimensions. Subsequently, a cross-attention-based feature fusion module (CAFM) is designed to efficiently integrate spatio-temporal multimodal features. Finally, these fused features are incorporated into an enhanced ConvGRU network. For multi-step prediction, we adopt a Seq2Seq architecture combined with a multi-step rolling strategy. Prediction experiments for chlorophyll-a concentration in the eastern seas of China validate the effectiveness of the proposed framework. The results show that, compared to baseline models, OKG-ConvGRU exhibits significant advantages in prediction accuracy, long-term stability, data utilization efficiency, and robustness. This study provides a scientific foundation and technical support for the precise monitoring and sustainable development of marine ecological environments.

Keywords: ocean element; spatio-temporal prediction; ConvGRU; cross-attention; time-series remote sensing image

check for updates

Academic Editor: Martin Gade

Received: 10 June 2025
Revised: 18 July 2025
Accepted: 31 July 2025
Published: 2 August 2025

Citation: Xiao, R.; Chen, Y.; Miao, L.; Jiang, J.; Zhang, D.; Su, Z. OKG-ConvGRU: A Domain Knowledge-Guided Remote Sensing Prediction Framework for Ocean Elements. *Remote Sens.* **2025**, *17*, 2679. https://doi.org/10.3390/rs17152679

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

As a typical nonlinear dynamic system, marine ecosystems are regulated by the synergistic effects of multiple environmental factors—including physical, chemical, and biological processes—and exhibit highly complex spatio-temporal heterogeneity [1,2]. With increasing human activities, environmental issues such as offshore eutrophication and water pollution have continued to worsen. Consequently, high-precision and large-scale prediction for key ocean elements (e.g., chlorophyll-a concentration, sea surface temperature) has become critical for marine ecological management and disaster early warning [3,4].

Current ocean remote sensing prediction methods primarily follow two paradigms: physics-driven and data-driven approaches [5,6]. Physics-driven methods, such as numerical models (e.g., ROMS, FVCOM), simulate ocean parameter evolution by solving simplified fluid dynamics equations derived from the Navier–Stokes equations [7]. While these models offer high interpretability, they face limitations including complex multi-physics field coupling, parameterization errors, and substantial computational costs, which restrict their practical application. In contrast, data-driven approaches have gained prominence with advances in artificial intelligence and the growing availability of ocean remote sensing data [8,9]. Researchers have successfully employed various machine learning models for ocean element prediction, including random forest (RF) [10], support vector machine (SVM) [11], and artificial neural network (ANN) [12,13]. Among these, convolutional neural networks (CNNs) excel at capturing spatial correlations in gridded data, leveraging convolutional operations to extract multi-scale features from ocean element images [14], while long short-term memory (LSTM) models demonstrate superior capability in modeling temporal dependencies within sequential data [15,16]. To address the need for joint spatio-temporal modeling, hybrid approaches such as CNN-LSTM and ConvLSTM have been developed. CNN-LSTM employs a serial architecture to separately handle spatial and temporal dynamics, though it may suffer from information loss between modules [17,18], whereas ConvLSTM integrates convolutional operations into LSTM gating units to simultaneously resolve spatio-temporal interactions, albeit with increased computational complexity [19,20]. Recent advancements, such as self-attentive mechanisms (SAM), have further enhanced feature representation by capturing long-range dependencies, though their black-box nature and sensitivity to data quality remain challenges [21,22]. Despite these advancements, data-driven models remain limited by their inability to fully explain complex ocean element interactions, ultimately affecting the reliability of the prediction results.

To address the limitations of existing prediction methods, researchers have begun to explore novel modeling approaches that integrate prior knowledge to enhance prediction accuracy while increasing model interpretability [23]. Among these, knowledge graphs (KGs) have gained increasing attention due to their structured representations and symbolic reasoning capabilities [24,25]. KGs store vast amounts of factual knowledge in triplets (head entity, relation, tail entity), effectively establishing domain-specific knowledge systems through interlinked entities and relations. In geospatial information science, scholars have developed specialized geographic knowledge graphs, such as GeoKGs [26,27], GIS KGs [28], flood KGs [29], UrbanKG [30], and RSKG [31]. These knowledge graphs not only model domain-specific knowledge but also support downstream applications by providing structured and interpretable representations. For instance, GeoKGs offer a novel framework for understanding, representing, and mining geoscientific knowledge through the integration of Earth big data, geoscientific knowledge, and models [26]. UrbanKG demonstrates remarkable performance in urban functional identification by integrating multi-source urban data [30,32,33]. Furthermore, the context-aware knowledge graph (CKG)-based traffic flow prediction model significantly enhances prediction accuracy by capturing the complex relationships of urban spatial and temporal contexts [34].

However, despite significant advancements in the implementation of KGs across various domains, the integration of domain knowledge in oceanic elemental Earth observations confronts distinct technical challenges. Firstly, in contrast to geographic features, ocean elements manifest intricate nonlinear spatial and temporal evolution patterns [1,2]. This necessitates the establishment of ocean KGs capable of capturing the spatial and temporal dependencies of ocean elements, processing interaction patterns, and undergoing dynamic updates. Secondly, the integration of marine domain knowledge with remote sensing data presents significant technical barriers in multimodal feature learning. Current KG-

based approaches generally encounter challenges in extracting semantic features from marine knowledge, primarily due to the absence of standardized ontologies and the intricacy of marine ecosystem relationships [35,36]. Furthermore, the alignment and fusion of these semantic features with high-dimensional visual features from time-series remote sensing imagery poses significant technical challenges, frequently leading to information loss or feature misalignment [31]. Thirdly, the integration of fused features into conventional prediction networks is constrained by architectural limitations. The majority of existing spatio-temporal prediction models have not been designed to accommodate graph-structured knowledge, leading to the suboptimal integration of domain knowledge [37]. This necessitates the development of innovative network architectures capable of effectively leveraging structured knowledge representations and conventional spatio-temporal data models. To our knowledge, no one has yet used knowledge graph techniques to predict typical ocean elements (e.g., chlorophyll-a concentrations).

To address the aforementioned challenges, this study proposes a domain knowledge-guided remote sensing prediction framework for ocean elements (OKG-ConvGRU). The framework consists of four core modules: the ocean elements spatio-temporal knowledge graph (OKG), semantic representation of the knowledge graph, a cross-attention-based multimodal feature fusion module (CAFM), and spatio-temporal prediction with an enhanced ConvGRU network. Specifically, the OKG is first constructed based on the domain knowledge of ocean elements, followed by semantic embedding representations for its spatial and temporal dimensions. Subsequently, CAFM is designed to deeply integrate the semantic features of the OKG with time-series remote sensing image features. Finally, these fused features are integrated into the enhanced ConvGRU network. For long-term prediction, a strategy combining the Seq2Seq architecture with multi-stage rolling prediction is adopted to enhance prediction stability. Compared to traditional knowledge-guided methods, OKG-ConvGRU demonstrates unique advantages in marine element prediction across multiple dimensions. Firstly, it employs structured modeling of the geographic spatial distribution, temporal variations, and influencing mechanisms of marine elements, which can effectively represent the complex interactions among physical, chemical, and biological factors. This significantly improves the explainability of the model. Secondly, the spatio-temporal knowledge of marine elements and time-series remote sensing data are efficiently fused by introducing a cross-attention-based feature fusion module (CAFM). Subsequently, the fused spatio-temporal features are learned using the enhanced ConvGRU network, which significantly improves the prediction accuracy of marine elements and the data utilization efficiency. In addition, by integrating Seq2Seq architecture with multi-stage rolling prediction, OKG-ConvGRU significantly improves the stability of long-term forecasting. The contributions of this study can be summarized in the following three aspects:

(1) A spatio-temporal knowledge graph of ocean elements (OKG) is constructed, effectively representing the geospatial distribution characteristics, temporal change patterns, and influence mechanisms of key ocean elements in a structured manner.
(2) A domain knowledge-guided remote sensing prediction framework for ocean elements is proposed, which combines the knowledge graph with the ConvGRU network for the first time. Based on the cross-attention mechanism, CAFM effectively fuses the spatio-temporal semantic features in OKG and the visual features of time-series remote sensing images, thereby enhancing the model's prediction performance for ocean elements.
(3) The performance of the OKG-ConvGRU-based chlorophyll-a concentration prediction model is evaluated using the eastern seas of China (Bohai Sea, Yellow Sea, and East China Sea) as an experimental area. The experimental results show that, compared with the baseline model, the proposed model exhibits significant advantages in prediction accuracy, long-term prediction stability, data utilization efficiency, and robustness.

2. Materials

2.1. Study Area

The geographical area under consideration is located in the eastern seas of China. The study area is situated within a broad range of 21° to 41°N latitude and 115° to 127°E longitude, as shown in Figure 1. This region encompasses three distinct sea areas: the Bohai Sea, the Yellow Sea, and the East China Sea. The region is influenced by freshwater input from mainland China, the Korean Peninsula, and numerous rivers in Japan, resulting in substantial changes in marine water quality and complex environmental factors. In recent years, intensified human activities have led to a marked deterioration in the quality of the sea area, characterized by deepening eutrophication of the water body and frequent occurrences of harmful algal bloom events [38,39]. In this context, the monitoring and prediction of key marine parameters, such as sea surface chlorophyll-a concentration (Chl-a) and sea surface temperature (SST), are not only important for regional ecological assessment, but also provide a scientific basis for the sustainable development of China and East Asia, which has significant environmental and economic values [15,40].

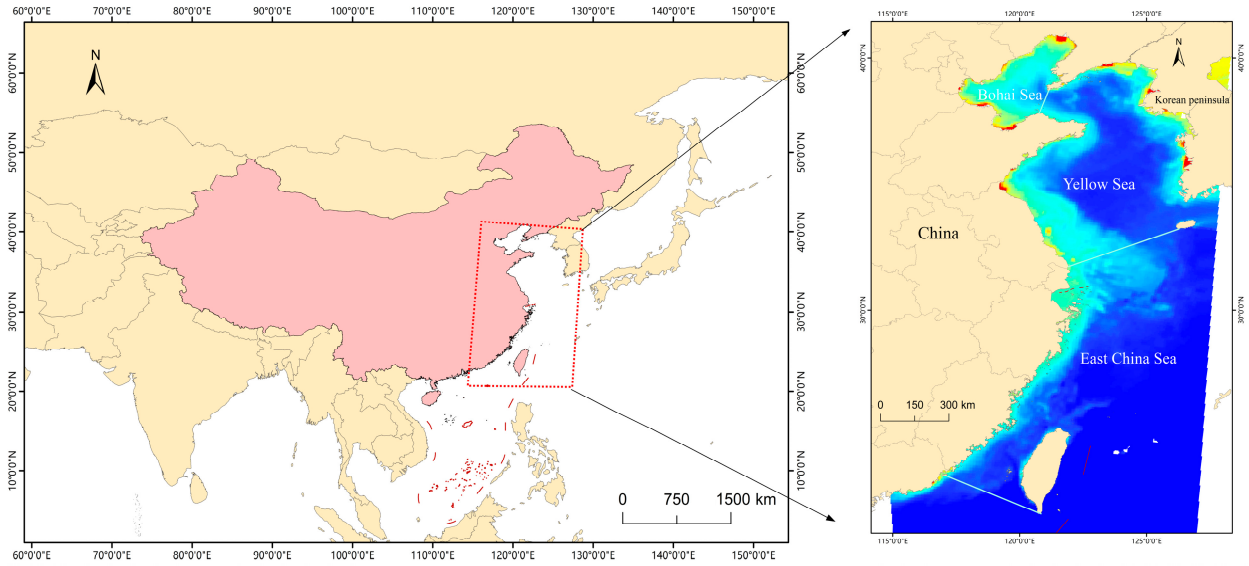


Figure 1. Geographic location of the study area based on Chl-a imagery from the MODIS L3 OceanColor product.

2.2. Raw Dataset

In this study, chlorophyll-a concentration (Chl-a) was selected as the target element for model prediction. Chl-a is influenced by sea surface temperature (SST), particulate inorganic carbon (PIC), particulate organic carbon (POC), photosynthetically active radiation (PAR), and normalized fluorescence line brightness (NFLH) [41]. According to existing studies, phytoplankton growth is affected by multiple interactions of physical, chemical, and biological factors [42,43]. Among these factors, SST shows a significant correlation with Chl-a concentration [44], while interactions among POC, PIC, and Chl-a reflect the productivity and carbon cycling processes in marine ecosystems [45,46]. In addition, PAR is strongly positively correlated with Chl-a [47,48].

The experimental data were obtained from satellite remote sensing images provided by NASA, spanning approximately 22 years, from August 2002 to May 2024, with a monthly temporal resolution. Monthly time-series data are monthly averages. The data were derived from the MODIS L3 OceanColor product, available through an open-access website (https://oceancolor.gsfc.nasa.gov/l3/ (accessed on 3 July 2024)), with a spatial resolution of 4 km. Detailed information is presented in Table 1.

Table 1. Datasets used for model training, validation, and testing.

Notation	Symbol Meaning	Unit	Spatial Resolution	Time Resolution	Data Sources
Chl-a	Chlorophyll-a concentration	mg/m^3	4 km × 4 km	Monthly	OceanColor
SST	sea surface temperature	degree_C			
PIC	Particulate inorganic carbon concentration	mol/m^3			
POC	Particulate organic carbon concentration	mol/m^3			
PAR	Photosynthetically active radiation	$einstein/m^2/day$			
NFLH	Normalized fluorescent line brightness	$W/m^2/um/sr$			

3. Methodology

3.1. Overview

For each type of marine environmental element in the study area, its remote sensing time-series image data can be represented as $\{X_t \mid t = 1, 2, \cdots, n\}$, where $X_t = (x_{t1}, x_{t2}, \cdots, x_{tm})$, and x_{ti} $(i = 1, 2, \cdots, m)$ denotes the observed values at the i-th spatial location at time t. Here, n represents the length of the time-series, and m represents the total number of pixels in a single image. The purpose of this study is to utilize the sequential image data of the target element and its influencing elements over the past T time steps $\{X_{t-T+1}, \cdots, X_{t-1}, X_t\}$ to predict the images of the target element for the next k time steps $\{X_{t+1}, \cdots, X_{t+k-1}, X_{t+k}\}$ by learning their spatio-temporal evolution patterns.

To address the above problems, this paper proposes a domain knowledge-guided remote sensing prediction framework for ocean elements, named OKG-ConvGRU, as shown in Figure 2. Firstly, we construct an ocean elements spatio-temporal knowledge graph (OKG) and then perform semantic embedding representations of its spatial and temporal dimensions. Subsequently, we design a cross-attention-based feature fusion module (CAFM) to effectively fuse spatio-temporal multimodal features. After that, the fused features are integrated into an enhanced ConvGRU network. Finally, the spatio-temporal multi-step prediction of ocean elements is achieved based on the OKG-ConvGRU framework.

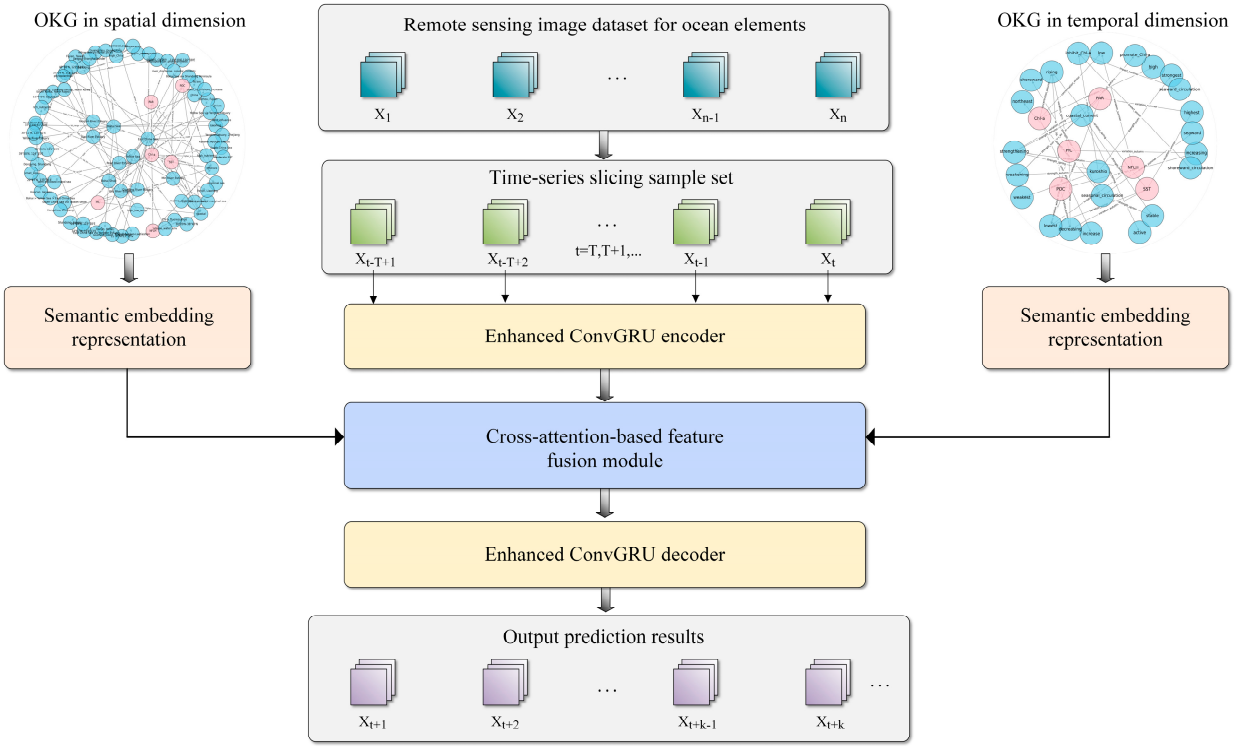


Figure 2. Overall architecture of ocean element prediction.

3.2. Construction of Spatio-Temporal Knowledge Graph for Ocean Elements

When processing remote sensing data of ocean elements, considering their close association with spatio-temporal characteristics, we categorize the relationships in the knowledge graph into spatial relationships (entity–relationship–entity), temporal relationships (entity–temporal relationship–time), and attribute relationships (entity–attribute–attribute value). This categorization facilitates comprehensive modeling of the spatial distribution characteristics, temporal variation patterns, and influencing mechanisms of ocean elements from multiple dimensions.

Existing spatio-temporal knowledge graphs mostly focus on urban areas with complex feature types, making them difficult to directly apply to marine scenarios due to significant differences in the elemental characteristics between urban and ocean environments. Therefore, we construct an ocean elements spatio-temporal knowledge graph (OKG) containing a total of 146 triplets, as shown in Figure 3. A list of all triples is shown in Table A1 in Appendix A. This graph encompasses both spatial and temporal dimensions, structurally representing the spatial distribution and temporal variation patterns of key ocean elements (Chl-a, SST, PIC, POC, PAR, NFLH) and related environmental factors within the study area. This provides a foundation for further integration of domain knowledge into the prediction model.

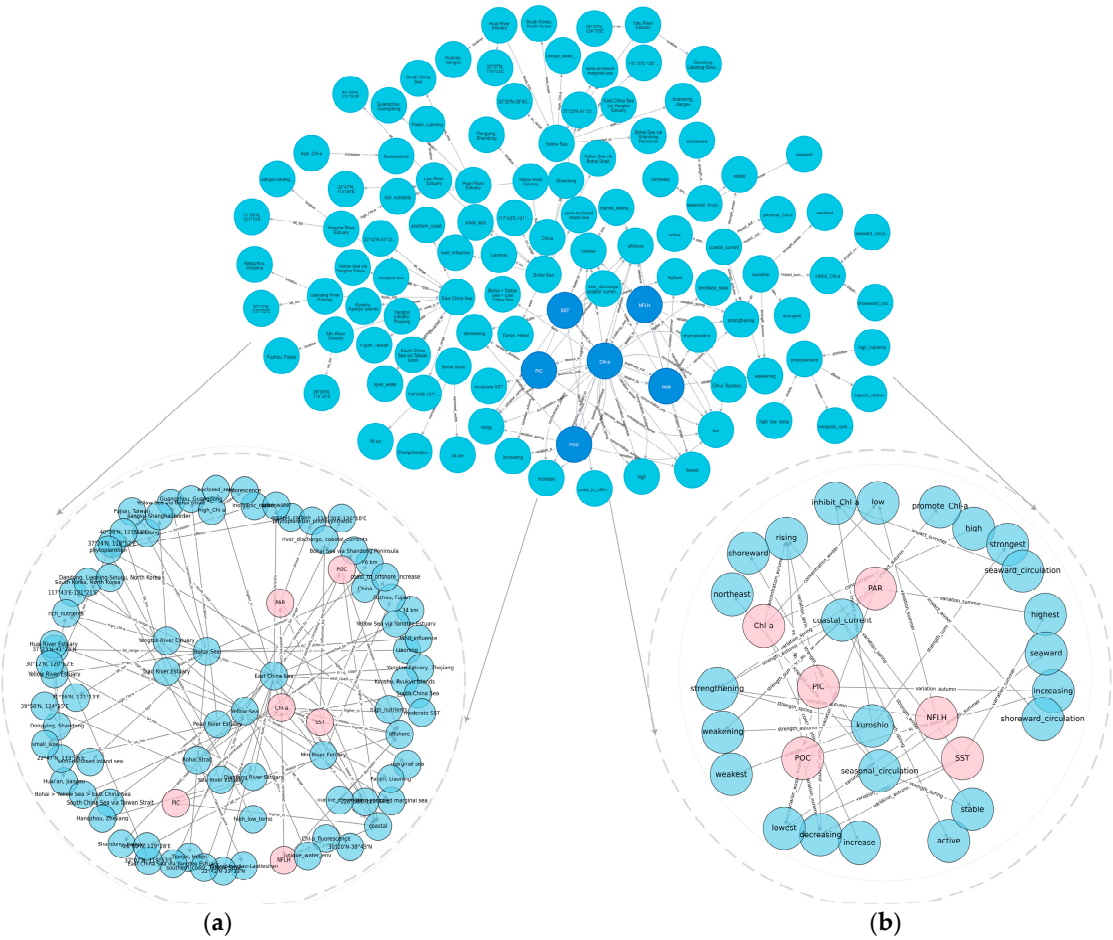


Figure 3. Spatio-temporal knowledge graph of ocean elements. (**a**) is its spatial dimension and (**b**) is its temporal dimension. Among them, red represents marine entities, and blue represents other entities.

3.2.1. Construction of Knowledge Graph in Spatial Dimension

The spatial dimension of OKG is designed to reveal the spatial distribution patterns of ocean, land, river inlets, and key ocean elements in the remote sensing images of the eastern seas of China, as well as their mutual influence mechanisms; it contains a total of 93 triplets (Figure 3a). The specific components are as follows:

(1) Spatial distribution of sea area and land area

The eastern seas of China include the Bohai Sea, the Yellow Sea, and the East China Sea. First, the latitude and longitude range of each sea area is defined (sea area-latitude/longitude range-degree range), for example, (Bohai Sea, latitude_range_, 37°23′N−41°23′N). Second, the spatial relationship between each sea area and the adjacent land is defined (sea area-spatial relationship-land/sea area) to clearly illustrate the spatial pattern of the sea area and the land.

(2) Spatial distribution of estuaries

The river estuary is an important node connecting land and sea, significantly impacting the distribution of ocean elements. In this study, we selected the major river estuaries flowing into the aforementioned sea area, including the mouths of the Yellow River, Liao River, Yalu River, Huai River, Yi River, Yangtze River, Qiantang River, Min River, and Pearl River. For each estuary, we describe its geographic location (estuary-latitude/longitude-degrees), for example, (Yellow River Estuary, latitude, 37°24′N), along with its administrative location (estuary-located in-province/city) and its inflow to the sea (estuary-inflow-sea area).

(3) Spatial distribution pattern of ocean elements

For the six major ocean elements observed by remote sensing, we describe their spatial distribution in different sea areas (sea area–ocean elements–characteristics), coastal characteristics (ocean elements–characteristics–coast), offshore gradient changes (ocean elements–change characteristics–offshore), estuarine distribution patterns (estuaries–ocean elements–characteristics), and the impact of estuaries on their distribution (estuaries–effects–ocean elements).

(4) Influence mechanism between major ocean elements

There are complex physical, biological, and chemical interactions among ocean elements. To systematically summarize these mechanisms and laws, we extensively collected research data, with Chl-a as the core research element, and sorted out its interactions with other elements (Chl-a–relationship–other elements), as well as the relationships among other elements (other ocean elements–relationship–other elements).

3.2.2. Construction of Knowledge Graph in Temporal Dimension

The long time-series remote sensing images of ocean elements exhibit significant periodic change characteristics. The temporal dimension of OKG is designed to describe this pattern of change over time and contains a total of 53 triplets (Figure 3b). We labeled each input image with a time indicator, classified by season, spring (March–May), summer (June–August), fall (September–November), and winter (December–February), and described the elements that are sensitive to temporal variations. The specific classifications are as follows:

(1) Seasonal change patterns of ocean elements

We analyzed the changing patterns of values and characteristics of each ocean element across the four seasons (ocean elements–seasons–characteristics) to understand their dynamic change throughout the year.

(2) Seasonal change rules of ocean currents

Ocean currents are the main driving force for the transportation and mixing of ocean elements, and their temporal changes are crucial for understanding the cyclic pattern of ocean elements. In this study, we select three ocean currents that play an important role in the eastern seas of China: the Kuroshio Current, the Littoral Current, and the Seasonal Circulation. We describe their different characteristics (current–seasonal–characteristics) and changes in their area of influence (current–seasonal–area of influence) over time.

(3) Mechanisms of ocean currents affecting ocean elements

Seasonal changes in ocean currents drive variations in ocean temperature, salinity, and nutrients, which in turn lead to cyclical patterns of ocean elements. This study details this influence mechanism (ocean currents–influence mechanism–ocean elements) and analyzes the temporal correlation between ocean currents and ocean elements in depth.

3.3. Semantic Embedding Representation of Knowledge Graphs

To effectively integrate domain knowledge from knowledge graphs into deep learning-based prediction models, this study adopts a knowledge graph embedding technique. Knowledge graph embedding projects the symbolic representation of a knowledge graph onto a low dimensional vector space, achieving a numerical representation of its semantic information. This allows entities and relationships with similar semantics to be closer together in the vector space, providing a foundation for downstream knowledge-guided machine learning tasks.

Inspired by the phenomenon that word vectors are translation invariant in semantic space, Bordes et al. (2013) [49] proposed the classical representation learning model TransE. For a triple (h, r, t), the model assumes that the vector representation of the head entity h plus the vector representation of the relation r should be equal to the vector representation of the tail entity t:

$$h + r \approx t \tag{1}$$

By minimizing the distance error of the triples in the embedding space, TransE learns the vector representations of entities and relations, and thus effectively predicts the missing links in the knowledge graph. The advantages of TransE lie in its computational efficiency, simplicity of implementation, and its excellent performance in dealing with simple relations. However, TransE's uniform treatment of embeddings for entities and relations leads to suboptimal performance when dealing with complex relations such as one-to-many, many-to-one, and many-to-many relations.

To address this limitation, an enhanced representation learning model, TransH (Wang et al. 2014) [50], has been proposed. This model introduces the concept of a hyperplane, which posits that each relation can be represented by a hyperplane on which translation operations are performed. Specifically, for a triple (h, r, t), TransH first projects the head entity h and the tail entity t onto the hyperplane corresponding to the relation r. Then, it performs a translation operation between the two projection vectors.

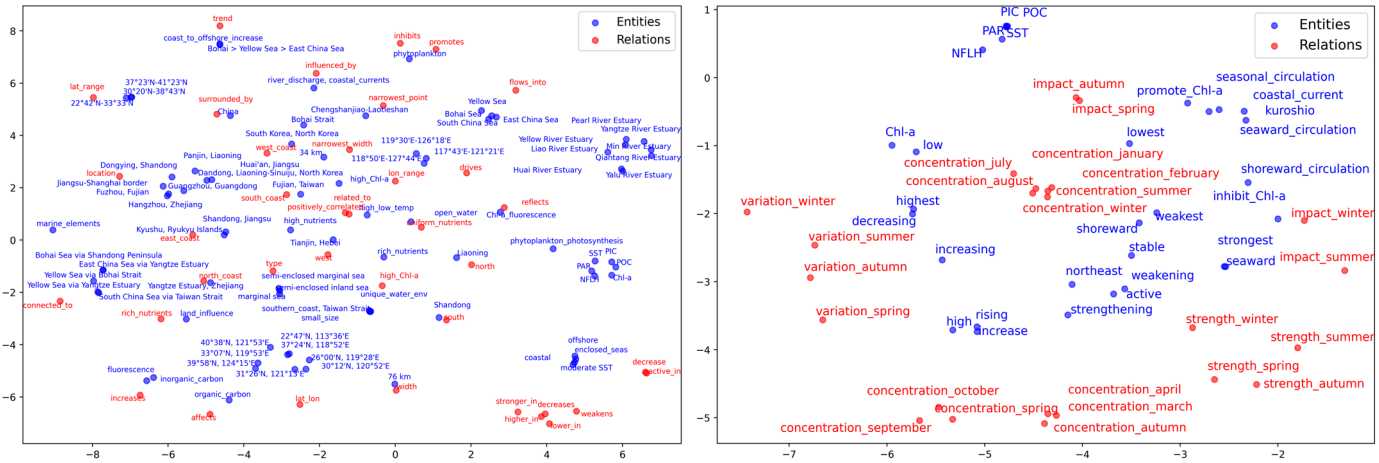
$$h_r + r \approx t_r \tag{2}$$

where h_r, t_r are the projection vectors of the head entity h and the tail entity t on the relation r hyperplane. In this way, TransH is able to complex relationships more effectively; however, this improvement also increases the number of parameters and the computational cost of the model, leading to relatively slow training and inference.

The inputs to the semantic embedding representation module in the framework are all the triples in the established OKG to obtain the semantic feature vectors of the

ocean elements in the spatial and temporal dimensions, which cover the spatio-temporal features of the six ocean elements (Chl-a, SST, PIC, POC, PAR, NFLH) in the four seasons. Since the knowledge graph constructed in this paper contains relatively simple types of relationships in time and space dimensions, the translation assumption of TransE is more suitable for dealing with such simple relationship patterns, while its efficient computational performance and easy implementation characteristics are more in line with the research needs. Therefore, in this paper, TransE is chosen as the embedding model, while TransH is used as the reference model in the evaluation to verify the applicability and advantages of TransE in this task.

To visualize the effect of knowledge graph embedding, we use the T-distributed stochastic neighborhood embedding (T-SNE) dimensionality reduction method to visualize the distribution of entities and relations in spatial and temporal dimensions in semantic space, as shown in Figure 4.



(**a**) (**b**)

Figure 4. Feature vectors after OKG embedding. (**a**,**b**) show the distribution of entities and relationships in the spatial and temporal dimensions, respectively.

3.4. Multimodal Feature Fusion Based on Cross-Attention Mechanism

The cross-attention mechanism is a variant of the attention mechanism. Its core idea is to compute the similarity between the query vector and the key vector, i.e., the attention scores, by using the feature vectors of one modality as the query vector (Query) and the feature vectors of another modality as the key vector (Key) and the value vector (Value). The value vectors are then weighted and aggregated based on these values to generate a new feature representation [51]. In recent years, the cross-attention mechanism has shown great potential in the field of feature fusion and has been successfully applied to tasks such as image–sentence matching [52], image fusion [53], and multispectral target detection [54], with notable results.

In this phase, our objective is to fuse the semantic feature vectors represented by the knowledge graph embedding with the image feature vectors extracted by the ConvGRU encoder, so that the image features can be adjusted and optimized with targeted guidance from domain knowledge and ultimately generate a fused feature vector suitable for the ConvGRU decoder. For this purpose, we designed a cross-attention fusion module (CAFM), which consists of a pair of multimodal information fusion modules (MIFMs) and a spatio-temporal information integration module (SIIM), as shown in Figure 5. Specifically, the semantic feature vectors of spatial and temporal dimensions are separately fused with the image feature vectors in MIFM. Subsequently, the two fused features generated are further

integrated in SIIM to obtain the final fused feature that includes both temporal and spatial characteristics. The process of feature fusion can be formalized as follows:

$$F_f = \mathrm{CAFM}(F_i, F_s, F_t) \tag{3}$$

where F_i represents the image feature vector before fusion, F_s, F_t represents the spatial and temporal semantic feature vectors, and F_f represents the fused image feature vector, whose magnitude is the same as that of F_i.

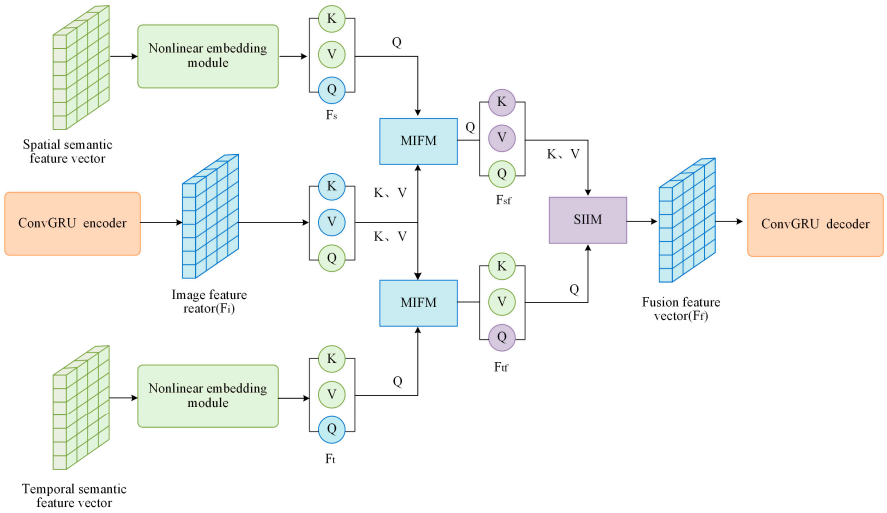


Figure 5. Cross-attention fusion module (CAFM) architecture.

Furthermore, before performing MIFM, a learnable nonlinear embedding module (NEM) is designed to reduce modal differences in input features, as shown in Figure 6. NEM projects the semantic feature representations of the knowledge graph onto a space shared with the visual features of the image to achieve their semantic alignment. It consists of two fully connected (FC) layers and a Gaussian error linear unit (GELU) activation function [32]. This architecture significantly improves the performance of the model in handling complex multimodal data by enhancing its nonlinear capability.

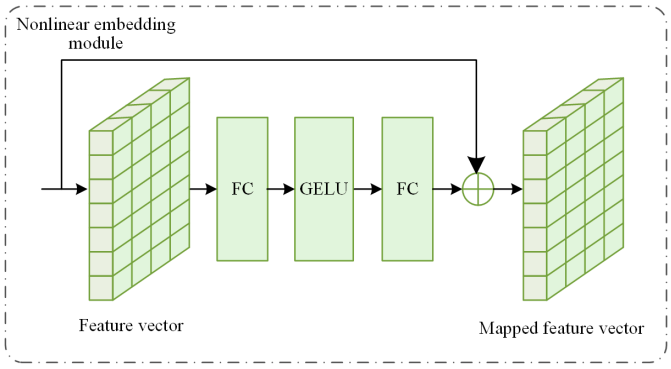


Figure 6. Nonlinear embedding module.

3.4.1. Multimodal Information Fusion Module (MIFM)

In this module, we optimize image features based on the multi-head cross-attention mechanism and domain knowledge contained in the semantic vectors, ultimately obtaining adjusted visual–semantic fused features, as shown in Figure 7. The process can be represented as follows:

$$F_{sf} = MIFM(F_i, F_s) \tag{4}$$

$$F_{tf} = MIFM(F_i, F_t) \tag{5}$$

where F_{sf}, F_{tf} represent visual–semantic fused feature vectors in spatial and temporal dimensions, respectively. Specifically, semantic features are used as query vectors and image features are used as key vectors and value vectors. The realization steps are as follows:

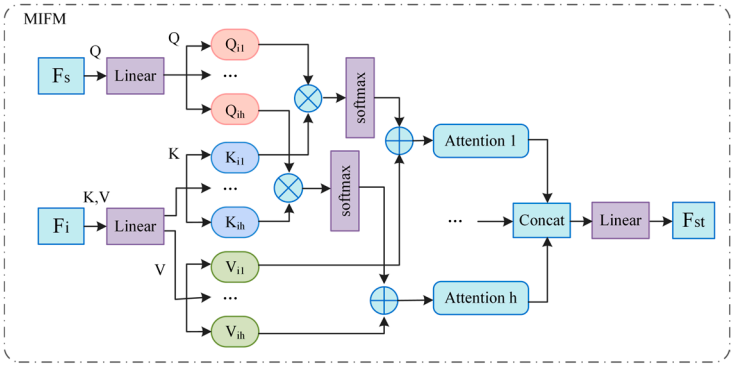


Figure 7. Multimodal information fusion module (MIFM).

First, each modal feature is divided into multiple parts, and multiple query, key, and value vectors are generated by linear transformation:

$$Q_{sh} = F_s W_h^Q,\ K_{ih} = F_i W_h^k,\ V_{ih} = F_i W_h^V\ \ (h = 1, 2, 3, \ldots, h) \tag{6}$$

where h represents the number of heads; W_h^Q, W_h^k, W_h^V are the weight matrices of the query, key, and value of the h-th head, respectively; and Q_{sh}, K_{ih}, V_{ih} are the vector matrices of the query, key, and value of the h-th header, respectively.

Subsequently, the dot product similarity between the query vector and the key vector is computed separately for each head. Then, the SoftMax operation is applied, and a weighted summation is performed with the corresponding value vector to derive the attention scores for each head, as expressed by the following formula:

$$\text{Attention}_h(Q_{sh}, K_{ih}, V_{ih}) = softmax(\frac{Q_{sh}K_{ih}^T}{\sqrt{D_k}})V_{ih}\ (h = 1, 2, 3, \ldots, h) \tag{7}$$

Finally, the outputs of all the heads are concatenated. Then, the feature dimensions of the outputs are made consistent with the inputs by linearly varying the formula as follows:

$$F_{sf} = Concat(Attention_1, Attention_2, \ldots, Attention_h)W_O \tag{8}$$

where W_O is a linear transform weight matrix with uniform feature dimensions. Equation (5) follows similar steps as Equation (4), with the difference that the spatial semantic feature vectors are replaced by temporal semantic feature vectors, where the former are fixed as inputs during the model training phase, while the latter are continuously adapted for fusion of the corresponding features based on the corresponding time points of the images.

3.4.2. Spatio-Temporal Information Integration Module (SIIM)

The purpose of this module is to integrate the multimodal fusion features obtained from the previous two MIFMs, so as to output integrated feature information in both temporal and spatial dimensions. As shown in Figure 8, the process can be represented as follows:

$$F_f = \text{SIIM}\left(F_{sf}, F_{tf}\right) \tag{9}$$

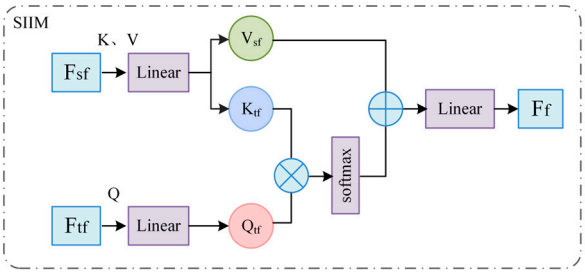


Figure 8. Spatio-temporal information integration module (SIIM).

To associate important feature information in the temporal dimension with that in the spatial dimension, we use temporally fused features as query vectors and spatially fused features as key vectors and value vectors and input them into SIIM for cross-attention-based integration.

Firstly, similar to the previous steps, the query vector of F_{tf} and the key and value vectors of F_{sf} and F_{tf} are transformed linearly into Q_{tf}, K_{sf}, V_{sf}, respectively. Then, a dot product attention layer is employed to calculate the similarity matrix between Q_{tf} and K_{sf}. This is followed by *softmax* operation and weighted summation with V_{sf} to obtain the interaction information between F_{sf} and F_{tf}. The process can be represented as follows:

$$Attention_f(Q_{tf}, K_{sf}, V_{sf}) = softmax(\frac{Q_{tf}K_{sf}^T}{\sqrt{D_k}})V_{sf} \tag{10}$$

Finally, in order to enable the output features to serve as inputs for the subsequent ConvGRU decoder, a linear transformation is applied to the integrated result as follows:

$$F_f = Attention_f\left(Q_{tf}, K_{sf}, V_{sf}\right)W_f \tag{11}$$

where W_f is a linearly transformed weight matrix, which ensures that F_f and F_i have the same dimension.

3.5. Enhanced ConvGRU Network

ConvGRU, an advanced variant of the Gated Recurrent Unit (GRU), is specifically designed to handle spatio-temporal data by integrating convolutional operations into its gating mechanisms, thereby replacing conventional matrix multiplications. This architecture comprises three fundamental components: the reset gate, update gate, and candidate activation. The reset gate governs the extent to which historical information is discarded, the update gate modulates the assimilation of new information, and the candidate activation produces a provisional state based on the current input and the reset gate's output. Collectively, these components enable the network to effectively capture temporal dependencies while preserving the spatial integrity of the data, making ConvGRU particularly adept at modeling time-series image data of ocean elements, which inherently exhibit spatio-temporal dependencies [55]. In comparison to ConvLSTM, ConvGRU offers a more streamlined architecture by eliminating the output gate, resulting in a reduction in the number of parameters, accelerated training speeds, and diminished sample size requirements. These attributes render ConvGRU a more computationally efficient and resource-effective solution for the dataset employed in this study.

However, ordinary ConvGRUs have limitations in handling hierarchical and multi-scale spatio-temporal features, thus failing to effectively encode and decode the spatio-temporal information in fused features. To address these limitations, we design an enhanced ConvGRU network as shown in Figure 9. In the encoder component of this enhanced network, a hierarchical three-layer architecture is employed. Each layer integrates a down-

sampling convolutional layer and a ConvGRU cell. The down-sampling convolutional layer systematically reduces the spatial resolution of the input images through 2D convolutional operations, while simultaneously capturing localized spatial features. The ConvGRU cell utilizes a gating mechanism to model temporal dependencies and extract multi-scale spatio-temporal features, thus encoding the image sequences into high-dimensional feature representations. Specifically, the convolutional operations within ConvGRU preserve spatial features by capturing local patterns, while the gating mechanism (update and reset gates) dynamically controls the flow of information to retain temporal dependencies. This ensures that both spatial and temporal features are effectively integrated and represented in the encoded feature space.

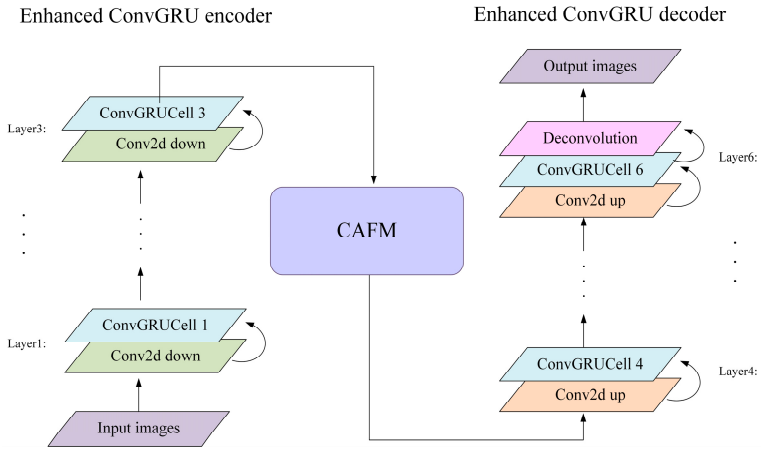


Figure 9. Enhanced ConvGRU network architecture.

In the decoder component, a symmetrical three-layer structure is adopted, with each layer consisting of an up-sampling layer and a ConvGRU cell. The up-sampling layer progressively restores the spatial resolution of the feature maps through interpolation or transposed convolution, ensuring the preservation of fine-grained details. The ConvGRU cell further refines the integration of spatio-temporal features to maintain temporal coherence in the predictive outputs. The final layer incorporates an inverse convolutional operation to enhance the precision of image details, resulting in high-resolution predictions. This encoder–decoder framework facilitates the effective extraction and reconstruction of spatio-temporal features of ocean elements, providing robust and discriminative feature representations for subsequent predictive modeling tasks.

3.6. Spatio-Temporal Multi-Step Prediction of Ocean Elements Based on OKG-ConvGRU

Traditional multi-step rolling prediction iteratively updates the dataset by adding the single-step prediction results to the end of the input sequence to achieve multi-step prediction [56,57]. This method is simple and easy to implement, with low computational complexity; however, it suffers from the cumulative error problem, where the prediction error gradually accumulates during the iteration process, resulting in decreased accuracy as the prediction steps increase. Additionally, this method relies solely on local information and may overlook global time dependencies.

To address these issues, this study introduces the Seq2Seq (Sequence-to-Sequence) architecture [58,59] and combines it with the concept of multi-step rolling prediction, proposing a multi-step prediction method that leverages the advantages of both approaches. The Seq2Seq architecture consists of two components: the encoder and the decoder. The encoder encodes the input sequences into a fixed-length vector, while the decoder generates the output sequences based on this vector. In the model design, both the encoder and

decoder adopt a four-layer OKG-ConvGRU stacking structure to effectively extract spatio-temporal features, as shown in Figure 10.

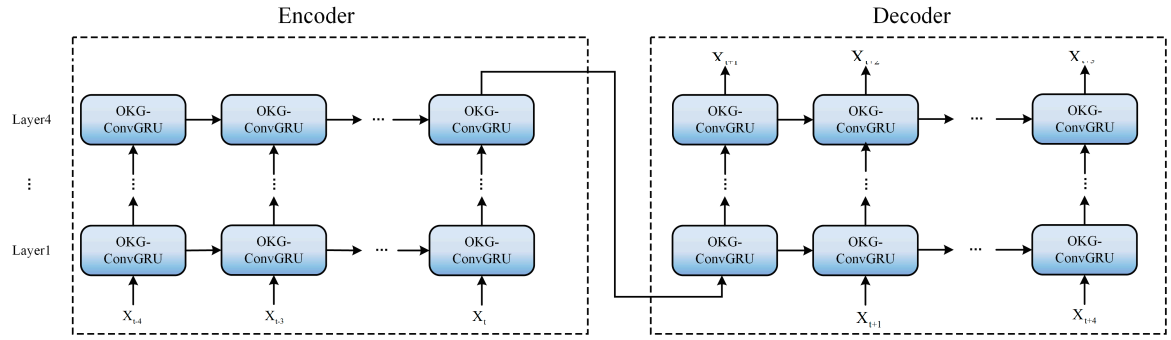


Figure 10. Multi-step prediction architecture.

During the training phase, the model learns the mapping relationship from the input sequence to the output sequence. When the number of prediction steps exceeds the decoder's step size, the multi-step rolling prediction approach is adopted, where the first time step data output from the decoder is fed back to the encoder as a new input sequence for iterative prediction. This approach not only retains Seq2Seq architecture's ability to model global dependencies but also enables flexible long-time sequence prediction through multi-step rolling, while reducing computational complexity. This method effectively addresses the cumulative error and local information dependence issues of traditional methods, enhancing the flexibility and adaptability of the model while ensuring prediction accuracy.

4. Experimental Results

4.1. Dataset Processing

4.1.1. Data Preprocessing

In this part, we performed several preprocessing operations on the original satellite images to improve the data quality and make them better adapt to the subsequent spatio-temporal prediction. To address the issue of missing values in original images, the data interpolation empirical orthogonal function (DINEOF) method [60,61] was utilized to reconstruct the missing image data. This method effectively restores the missing values and retains the spatio-temporal variation characteristics of the data through spatio-temporal covariance matrix decomposition and iterative interpolation. Subsequently, high-precision land vector data corresponding to the selected projection was employed to implement a masking process for the land anomalies of the ocean water color data, thereby eliminating geographic interference. We then processed the data for outliers, replacing negative values with 0 and using the Winsorization method to replace pixel values that exceeded the upper limit with the upper limit. To unify the dimensions of the multi-source data, the parameters were normalized to the [0, 1] interval by min–max normalization [62]. Finally, the images were uniformly cropped to 320 × 568 pixel specifications to fit the model inputs.

The dataset division strictly followed the principle of temporal continuity, and the 262 months of data from August 2002 to May 2024 (2002.08–2024.05) were divided into three subsets: the training set (2002.08–2018.05, 90 months) is used for model parameter learning, the validation set (2018.06–2021.05, 36 months) is used for hyperparameter optimization, and the test set (2021.06–2024.05, 36 months) is used to evaluate the model generalization ability.

4.1.2. Correlation Analysis

In order to verify the reasonableness of the input variables of our model, the nonparametric statistical method was used for multivariate correlation analysis. By calculating the Spearman correlation coefficient, the monotonic correlations between Chl-a and the other five marine environmental variables (i.e., SST, PIC, POC, PAR, NFLH) were quantified.

As demonstrated in Figure 11, the thermogram constructed based on Spearman's rank correlation coefficient (R) reveals the pattern of correlation between Chl-a and the other ocean elements. The intensity of the color scale is positively correlated with $|R|$. The results show that there were significant correlations between Chl-a and all five environmental variables. Specifically, the strongest and most significant negative correlation (R = -0.792, $p < 0.001$) is observed between Chl-a and SST, indicating that elevated water temperatures exert an inhibitory effect on algal metabolism. Conversely, positive correlations are identified between Chl-a and the other variables (PAR, POC, PIC, and NFLH), which reflect the positive effects of light and organic matter on phytoplankton growth. In addition, a significant cross-correlation is identified among the variables. For instance, a strong positive correlation is observed between PAR and SST (R = 0.863, $p < 0.001$), suggesting a synergistic effect between solar radiation and surface seawater thermodynamic processes. This multi-dimensional correlation network confirmed the ecological coupling of the input variables.

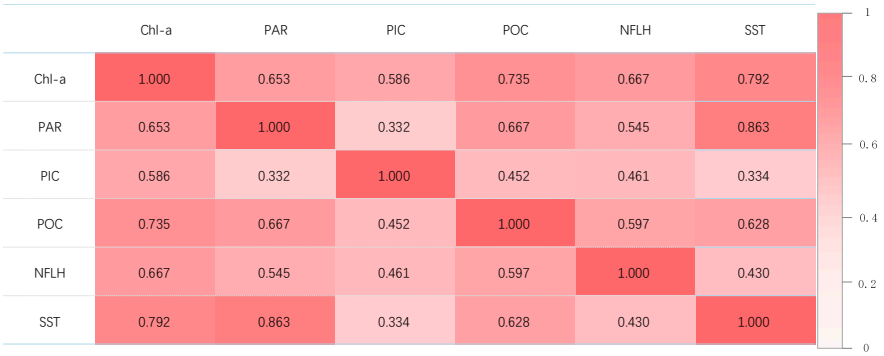


Figure 11. Heat map of Spearman's correlation coefficient intensity matrix.

4.1.3. Construction of Time-Series Slicing Sample Set

This study used a sequence of ocean element images from a continuous period of time in the past as input to predict their values in the future. To this end, we adopted a sliding window method along the timeline to slice the preprocessed time-series images and construct a sample dataset. As shown in Figure 12, each sample data consists of T time-series images, where the first T/2 images were used as inputs, and the observed values of the following T/2 images were used as their corresponding labels (i.e., predicted values).

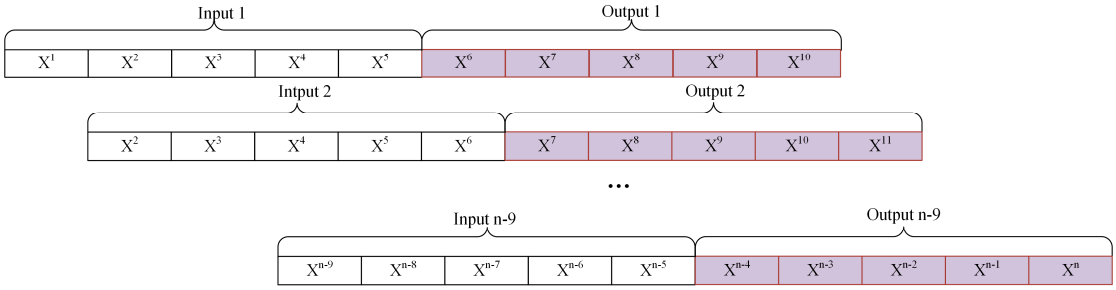


Figure 12. Time-series slicing sample construction.

Through extensive experiments and comparative analysis, we find that when the time-series slice length T is set to 10 months and the input and output sequence lengths are both set to 5 months, the prediction model can achieve excellent performance. When the predicted future time step k is greater than 5 months, adopting a multi-step rolling prediction strategy can achieve longer predictions. Specifically, using sequence image data of six ocean elements from the past five months, it is possible to predict Chl-a images for multiple time steps (k > five months) in the future.

4.2. Evaluation Metrics

In this study, three evaluation metrics were employed to quantitatively assess the overall prediction performance of the model: mean absolute error (MAE), root mean square error (RMSE), and goodness of fit (R^2). A lower MAE and RMSE indicate higher model prediction accuracy. R^2 is used to evaluate the degree of fit between the predicted values and observed values of the model, with values ranging from 0 to 1. When close to 1, the goodness of fit is high, indicating that the observed values are close to the expected values of the model, i.e., the difference between the model's predictions and the actual observations is small. Conversely, when close to 0, the model's predictions differ significantly from actual observations. The units of MAE and RMSE are the same as those of the predicted target element (Chl-a), mg/m^3. These indicators are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \tag{12}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i| \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}|x_i - y_i|^2}{\sum_{i=1}^{N}|x_i - \overline{x_i}|^2} \tag{14}$$

where N is the total number of samples, x_i is the actual observed value, y_i is the model's predicted value, and $\overline{x_i}$ is the average of the true values of all samples.

To evaluate the performance of the embedding model, we used three metrics: Mean Rank (MR), Mean Reciprocal Rank (MRR), and Hits at K (Hits@K). MR indicates the average position of the correct entity in the ranking of all predictions, with lower values indicating more accurate model predictions; MRR measures the average of the inverse of the rankings of the correct entities, which is more focused on whether the correct answer appears in the top ranks, and higher values are better; Hits@K reflects the proportion of correct entities included in the top K predictions and is used to assess the accuracy of the model, especially when K is small, and higher values indicate better model performance. Note that for a more accurate evaluation, we removed the real triples in training and calculated the filtered rankings to obtain the filtered metrics above. The specific formula is as follows:

$$MR\ \frac{1}{|T|}\sum_{t\in T} rank(t) \tag{15}$$

$$MRR\ \frac{1}{|T|}\sum_{t\in T} \frac{1}{rank(t)} \tag{16}$$

$$Hits@K\ \frac{1}{|T|}\sum_{t\in T} 1[rank(t) \leq K] \tag{17}$$

where T denotes the test set, *rank(t)* denotes the rank position of the correct entity (or relation) t in a particular query, and K represents that only the top K positions are considered when evaluating the prediction results.

4.3. Model Implementation Details

4.3.1. Experimental Environment

The experiments were conducted on a workstation that was equipped with an Intel Core i7-14650HX processor and operated on the Windows 11 operating system. The model was implemented based on the PyTorch framework (version 1.12) and utilized an NVIDIA RTX 4070 graphics card (16 GB video memory) for the purpose of training acceleration, with CUDA version 12.5. Code development and debugging were conducted in the PyCharm (version 2024.1.1) integrated development environment.

4.3.2. Model Settings

The model employed the mean square error (MSELoss) as the loss function, and the optimizer selected Adam's algorithm, whose learning rate was set to an initial value of 0.001, with a sampling variation rate of 0.00002 and the total number of training rounds set to 25,000. To ensure reproducibility and eliminate potential bias, the random seed was fixed throughout the training process. To address the challenge of computational resources posed by the increase in model complexity, this study employed a block-based training strategy. The input data was segmented into n sub-blocks (n is the number of blocks) along the channel dimensions. These sub-blocks were processed separately during the training process, and ultimately, the output results were integrated into the original image size at the prediction stage.

4.4. Knowledge Graph Embedding Evaluation

For the constructed OKG (which includes spatial and temporal dimensions), we used TransE and TransH models for embedding, respectively. In the training process of knowledge graph embedding models, the selection of hyperparameters significantly impacts model performance. To identify the optimal combination of hyperparameters, we employ a Bayesian optimization strategy to optimize the parameter configurations for both the TransE and TransH models. The objective function is defined as the Mean Rank (MR), aiming to minimize this metric. The hyperparameters to be optimized include the embedding dimension of entities and relations (ranging from 50 to 200, with a step size of 10), the margin parameter in the loss function (ranging from 0.5 to 2.0, with a step size of 0.1), the weight of the soft constraint (ranging from 0.01 to 0.5, with a step size of 0.01), the learning rate (ranging from 0.001 to 0.1, with a step size of 0.001), and the number of negative samples per positive sample (ranging from 1 to 50, with a step size of 1). In the Bayesian optimization process, we first construct a probabilistic model of the objective function using Gaussian Process (GP) and Expected Improvement (EI) as the acquisition function to guide the search in the parameter space. Through iterative evaluations, the objective function is assessed in each iteration, and the Gaussian Process model is updated accordingly, gradually approaching the global optimum until the objective function converges. The results show that TransE significantly outperformed TransH across all evaluation metrics under the optimal configuration (see Table 2). This result may be attributed to the fact that the types of relationships in the constructed knowledge graphs are relatively simple, and the TransE's translation assumption ($h + r \approx t$) is more suitable for handling such simple relationships. In contrast, TransH introduces relation-specific hyperplanes to model complex relations, but in this scenario, this complexity may be redundant and instead increase the model complexity and training difficulty.

Table 2. Effects of different embedding models.

Index	Spatial Dimension					Temporal Dimension				
	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
TransE	1.0914	0.9755	0.9570	0.9839	1.0000	1.0472	0.9796	0.9623	1.0000	1.0000
TransH	3.3019	0.4310	0.1321	0.6698	1.0000	3.3430	0.4521	0.1709	0.6815	0.9926

During the Bayesian optimization process, we observed that when both models achieved their optimal configurations (with the lowest MRR), TransE required significantly fewer negative samples, training epochs, batch sizes, and embedding dimensions compared to TransH. This difference stems from TransE's simpler structure, which has lower dependence on training resources and data while maintaining high performance even under resource constraints. In contrast, TransH's more complex architecture demands greater resources to prevent overfitting.

To further analyze the embedding model's performance, we visualized the relationships among the head entity, relation, and tail entity by randomly sampling triples and projecting them into a two-dimensional space using t-SNE (Figure 13). This was conducted for both spatial and temporal dimensions.

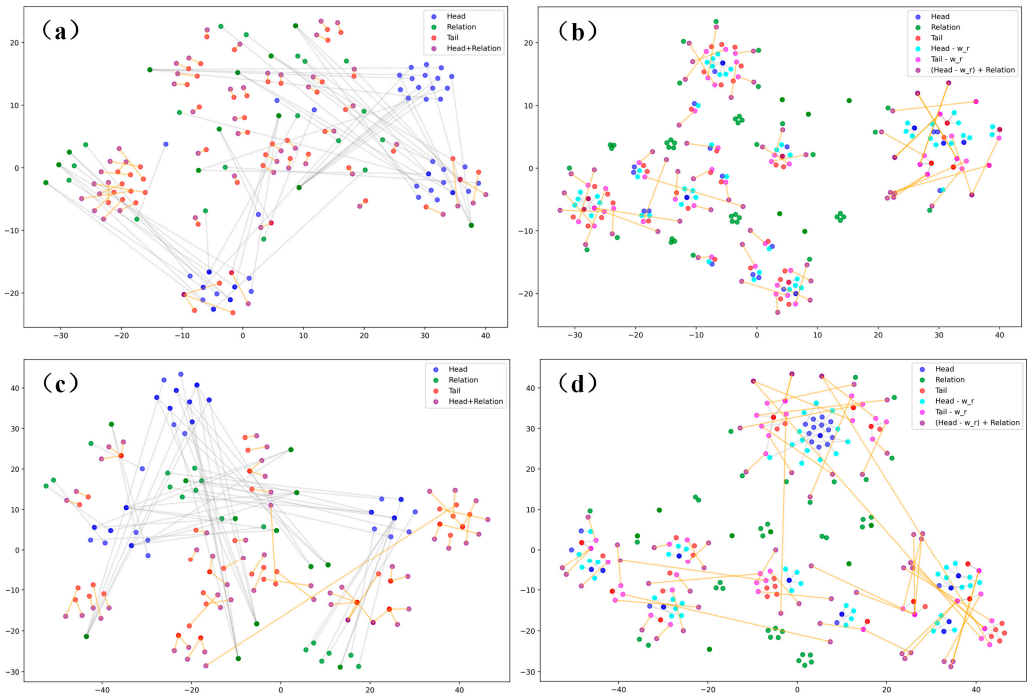


Figure 13. Distances between head entities, relations, and tail entities under different embedding models. Panels (**a**,**b**) present spatial dimension triple embeddings using TransE and TransH, respectively, while (**c**,**d**) show their temporal dimension counterparts.

Comparing Figure 13a,b with Figure 13c,d, the following phenomena were observed: (1) In TransE, the entity-to-relationship distance is larger, indicating that the model relies more on the relationship vectors to convey information rather than on entity similarity. Additionally, the distance from the head entity plus the relationship to the tail entity is smaller, which verifies the high efficiency of its tail entity prediction. (2) In contrast, in TransH, the head and tail entities after projection are concentrated in the pre-projections around the entities with smaller distances, indicating that the projection operation has limited effects on the entity vectors and the relation hyperplane fails to adequately capture the semantic changes in the entities. The larger distance from the projected head entity

plus the relation to the tail entity indicates that the relation vector fails to effectively model the semantic transformation, reflecting its inadequacy in capturing the ternary structure information. (3) The ternary distribution in TransE is more uniform, balancing the representation of entities and relations, whereas the distribution is more concentrated in TransH. The projection operation restricts the diversity of expression, resulting in the failure of the semantic information to unfold adequately.

The above analysis further confirms the significant advantages of TransE in embedding spatial and temporal knowledge graphs. To quantify this conclusion, we substituted each of the two embedding methods into the overall prediction model and calculated their R^2, MAE, and RMSE at the T + 1 time step during the test phase, as shown in Figure 14.

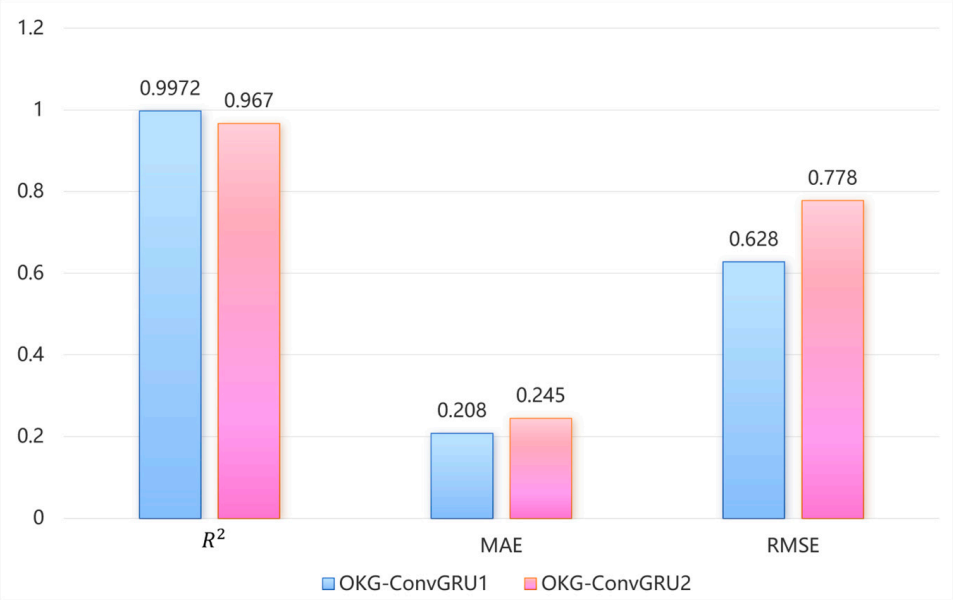


Figure 14. OKG-ConvGRU1 represents embeddings using TransE, while OKG-ConvGRU2 represents embeddings using TransH.

The experimental results show that the prediction performance of OKG-ConvGRU1 is significantly better than that of OKG-ConvGRU2, indicating that the quality of knowledge graph embedding has an important impact on the overall prediction accuracy of the model. Based on these results, we used TransE to embed spatial and temporal knowledge graphs in subsequent model applications to ensure the efficiency and accuracy of the model in prediction tasks.

4.5. Comparison of Predictive Performance with Benchmark Models

To comprehensively evaluate the performance of the proposed OKG-ConvGRU model in spatio-temporal oceanic element prediction, this study conducted a systematic experimental evaluation of Chl-a prediction. The evaluation was based on the test dataset of the study area spanning the period from June 2021 to May 2024, totaling 36 months. In the experimental design, the OKG-ConvGRU model was first compared with the single data-driven CA-ConvGRU (ConvGRU combined with a cross-attention mechanism) model. To further validate the model's superiority, we selected five types of current mainstream deep learning-based spatio-temporal prediction models as benchmark models, including the GRU model, the CNN-LSTM hybrid model, the ConvLSTM model, the ConvGRU model, and the SA-ConvLSTM (ConvLSTM with a self-attention module) model. In addition, we also used a simple forecasting model based on meteorology as a baseline model. This model is used as a predictor for the corresponding month in the test set by calculating

the historical average of the chlorophyll-a concentration data for the same month in the training and validation sets.

To ensure the scientific validity and reliability of the experimental results, all models were tested under the same experimental environment and dataset, using the same evaluation metrics. Table 3 shows the best values of each evaluation metric for Chl-a prediction by each model at five prediction time steps from T + 1 to T + 5, providing a reliable basis for quantitative comparison of model performance.

Table 3. Comparison of predictive performance of different models for Chl-a.

Model	Metric	T + 1	T + 2	T + 3	T + 4	T + 5
Climatological Mean Prediction	MAE	0.833	0.851	0.874	0.912	0.933
	RMSE	1.572	1.865	2.267	2.680	3.035
GRU	R^2	0.9536	0.9521	0.9516	0.9510	0.9504
	MAE	0.261	0.263	0.269	0.275	0.282
	RMSE	0.826	0.838	0.846	0.853	0.861
CNN-LSTM	R^2	0.9685	0.9672	0.9665	0.9658	0.9642
	MAE	0.236	0.242	0.247	0.255	0.265
	RMSE	0.763	0.776	0.780	0.784	0.792
ConvLSTM	R^2	0.9683	0.9672	0.9665	0.9653	0.9640
	MAE	0.245	0.252	0.259	0.267	0.278
	RMSE	0.805	0.811	0.817	0.826	0.838
ConvGRU	R^2	0.9697	0.9693	0.9690	0.9686	0.9682
	MAE	0.237	0.240	0.244	0.248	0.255
	RMSE	0.741	0.747	0.756	0.762	0.770
SA-ConvLSTM	R^2	0.9924	0.9922	0.9920	0.9923	0.9918
	MAE	0.228	0.234	0.237	0.231	0.240
	RMSE	0.694	0.701	0.708	0.697	0.715
CA-ConvGRU	R^2	0.9953	0.9950	0.9954	0.9957	0.9952
	MAE	0.215	0.223	0.212	0.206	0.217
	RMSE	0.642	0.649	0.638	0.633	0.645
OKG-ConvGRU	R^2	0.9972	0.9971	0.9973	0.9974	0.9972
	MAE	0.208	0.210	0.205	0.202	0.208
	RMSE	0.628	0.630	0.624	0.617	0.628

Through the comparative analysis of the values in Table 3, it can be observed that our proposed OKG-ConvGRU model significantly outperforms other models in future multi-step prediction. Specifically, the MAE and RMSE values of OKG-ConvGRU stabilize within 0.210 and 0.630, respectively, with minimum values reaching 0.202 and 0.617, respectively, which are the lowest values among all models. Meanwhile, its R^2 value remains stable above 0.9971, with a maximum value of 0.9974, further verifying the superiority of OKG-ConvGRU in terms of fitting effect and prediction accuracy.

Compared to the single-modal CA-ConvGRU, OKG-ConvGRU significantly improves all the metrics in multi-step prediction, which fully demonstrates the effectiveness of incorporating prior knowledge from the knowledge graph into the prediction model. Additionally, when comparing and analyzing the benchmark models, we find that the prediction accuracies of the CA-ConvGRU, SA-ConvLSTM, ConvGRU, CNN-LSTM, ConvLSTM, and GRU models decrease in that order. The Climatological Mean Prediction model has the worst performance. This may be due to the fact that the higher the model complexity and ability to capture spatio-temporal features, the higher the prediction accuracy, whereas the climate mean prediction model relies only on historical averages, which are not able

to capture complex dynamic changes. Among them, CA-ConvGRU further enhances the prediction accuracy by introducing the cross-attention mechanism compared to the SA-ConvLSTM model, which incorporates the self-attention module, indicating a stronger ability to capture spatio-temporally dependent information. Meanwhile, ConvGRU is more streamlined compared to ConvLSTM, effectively alleviating the model complexity problem caused by the introduction of the attention mechanism. To assess the statistical significance of differences in predictive performance among different models, this study employed a one-way analysis of variance (One-way ANOVA) to test the significance of the mean absolute error (MAE) metrics across seven models. As illustrated in Figure 15, the test results indicate that, except for the differences between ConvGRU and ConvLSTM, as well as CNN-LSTM, which did not pass the significance test ($p > 0.05$) likely due to their similar architectural characteristics and feature extraction mechanisms, the differences in MAE among all other models were statistically significant ($p < 0.05$). Notably, the differences in MAE between the proposed OKG-ConvGRU model and all baseline models reached an extremely significant level ($p < 0.001$), demonstrating that, at a 99% confidence level, the predictive performance of OKG-ConvGRU is significantly superior to that of the other baseline models.

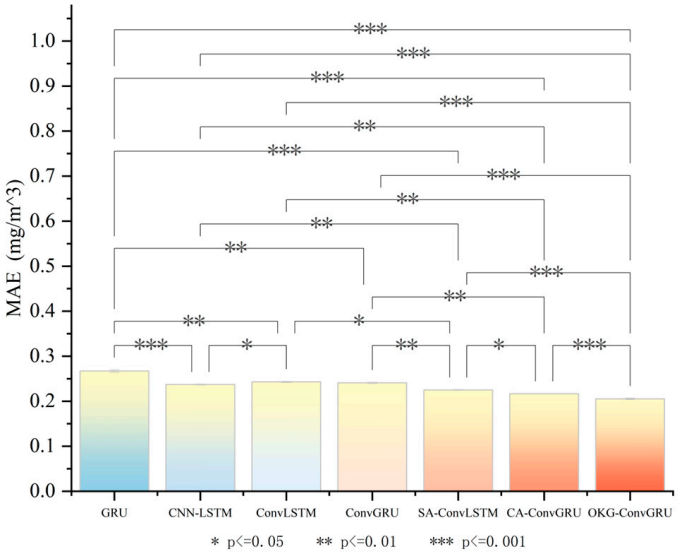


Figure 15. Significance test results for MAE of different models.

Further analysis of the performance of each model in multi-step prediction reveals that the prediction effectiveness of the ConvGRU, ConvLSTM, and CNN-LSTM models decreases significantly as the number of prediction steps increases. This error accumulation phenomenon mainly stems from the fact that each prediction step relies on the output of the previous step, resulting in an amplifying error. In contrast, the CA-ConvGRU, SA-ConvLSTM, and OKG-ConvGRU models exhibit different characteristics; their errors do not increase significantly with the number of prediction steps but instead show a fluctuating and relatively stable trend. This phenomenon may be attributed to the introduction of the attention mechanism, which allows the models to learn deeper patterns and long-term dependencies in long time-series. It is worth noting that the OKG-ConvGRU model demonstrates the strongest metric stability at all time steps, indicating that the prior knowledge in the knowledge graph effectively aligns the periodic spatio-temporal variation characteristics of the input images. This alignment helps the model to more accurately capture and maintain the intrinsic patterns and dynamic characteristics present in the sequence data.

To visually assess differences in model prediction performance across regions, we plotted the February 2024 (T + 1 time step) Chl-a concentration prediction results of multiple models in the study area, as shown in Figure 16. We also plotted the distribution of prediction errors at this time point, as shown in Figure 17. Comparing the prediction results and error distributions of the different models, we find that OKG-ConvGRU exhibits the best performance in terms of prediction accuracy and detailed feature capture, with its prediction trend highly consistent with the actual values. CA-ConvGRU tends to overestimate the Chl-a concentration along the Bohai Sea coastline, while the predictions from SA-ConvLSTM are relatively accurate in the Bohai Sea region but less reliable in the deep-sea region. This may be due to the fact that these two models rely too heavily on the attention mechanism and fail to fully capture the prevailing spatial and temporal dependencies and distribution patterns.

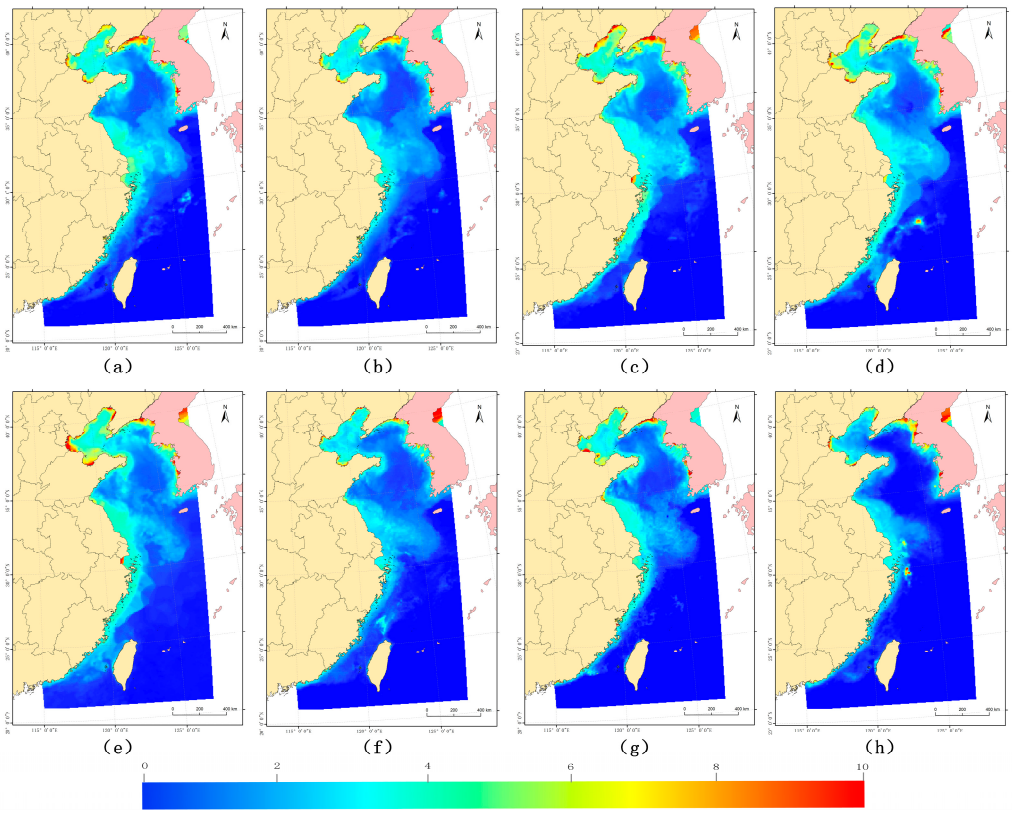


Figure 16. Prediction results of Chl-a for different models: (**a**) observed values of Chl-a; (**b**) OKG-CA-ConvGRU; (**c**) CA-ConvGRU; (**d**) SA-ConvLSTM; (**e**) ConvGRU; (**f**) ConvLSTM; (**g**) CNN-LSTM; (**h**) GRU.

In addition, the prediction results of ConvGRU, ConvLSTM, CNN-LSTM, and GRU are consistent with the actual values in terms of the overall trend. However, there are significant differences in the details, particularly in localized areas of the deep-sea. This phenomenon may be related to the limitations of these models in capturing complex spatio-temporal dependencies, especially their insufficient ability to model the nonlinear patterns of change in deep-sea regions.

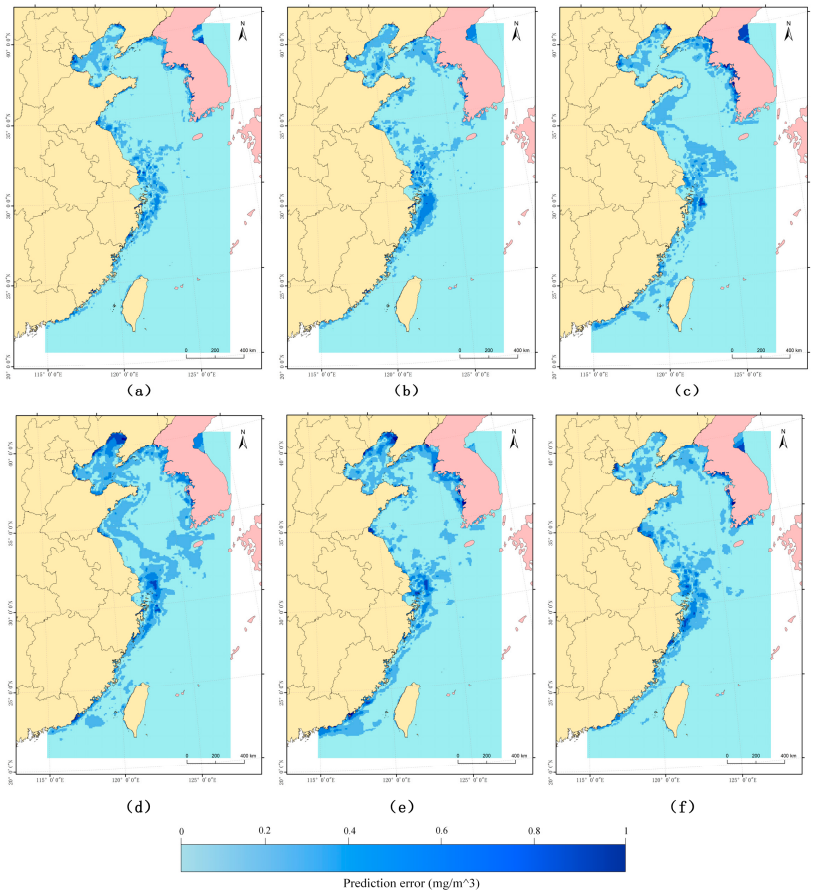


Figure 17. Distribution of the prediction errors for different models. The prediction error is obtained by subtracting the absolute value of the actual value from the predicted value. (**a**) OKG-CA-ConvGRU; (**b**) CA-ConvGRU; (**c**) SA-ConvLSTM; (**d**) ConvGRU; (**e**) ConvLSTM; (**f**) CNN-LSTM.

4.6. Long-Term Predictive Performance Evaluation

To further evaluate the performance of OKG-ConvGRU in multi-step prediction, we used the image data from five time steps in the test set (August 2022 to December 2022) as model inputs to predict Chl-a for the next ten time steps (January 2023 to October 2023). The long-term predictions of the model are shown in Figure 18.

By comparing the prediction results with the observations at different time steps, we find that the model demonstrates excellent performance in short-term prediction, with its prediction results highly consistent with the observations. However, as the number of prediction steps increases, a discrepancy between the predicted and observed values gradually emerges, which is especially significant in the deep-sea region far from the coast. Nevertheless, the model still exhibits good overall prediction performance. In the multi-step prediction task, this study innovatively adopted a hybrid prediction strategy that combines the Seq2Seq architecture with multi-step rolling prediction. To verify the effectiveness of this strategy, we designed a series of controlled experiments. With the premise of ensuring the consistency of the input data, we used both the hybrid prediction strategy and the traditional multi-step rolling prediction method to make predictions. The error metrics (MAE and RMSE) were calculated for different prediction step sizes, as shown in Figure 19.

Figure 19 shows the long-term prediction performance of the OKG-ConvGRU model under two prediction strategies. The results indicate that the error metrics (MAE and RMSE) increase significantly with longer prediction step lengths when using the multi-step rolling prediction approach, and the growth trend is nearly exponentially distributed, which aligns

with the pattern of error accumulation in iterative operations. In contrast, when the Seq2Seq architecture is combined with multi-step rolling prediction, the error metrics still exhibit an overall increasing trend, but the fluctuations are more stable and even decrease at certain time steps. Specifically, the MAE and RMSE of the model are stabilized within 0.225 and 0.658 for predictions made from January to October 2023, indicating that this strategy can effectively improve the accuracy and stability of the model in long-term prediction.

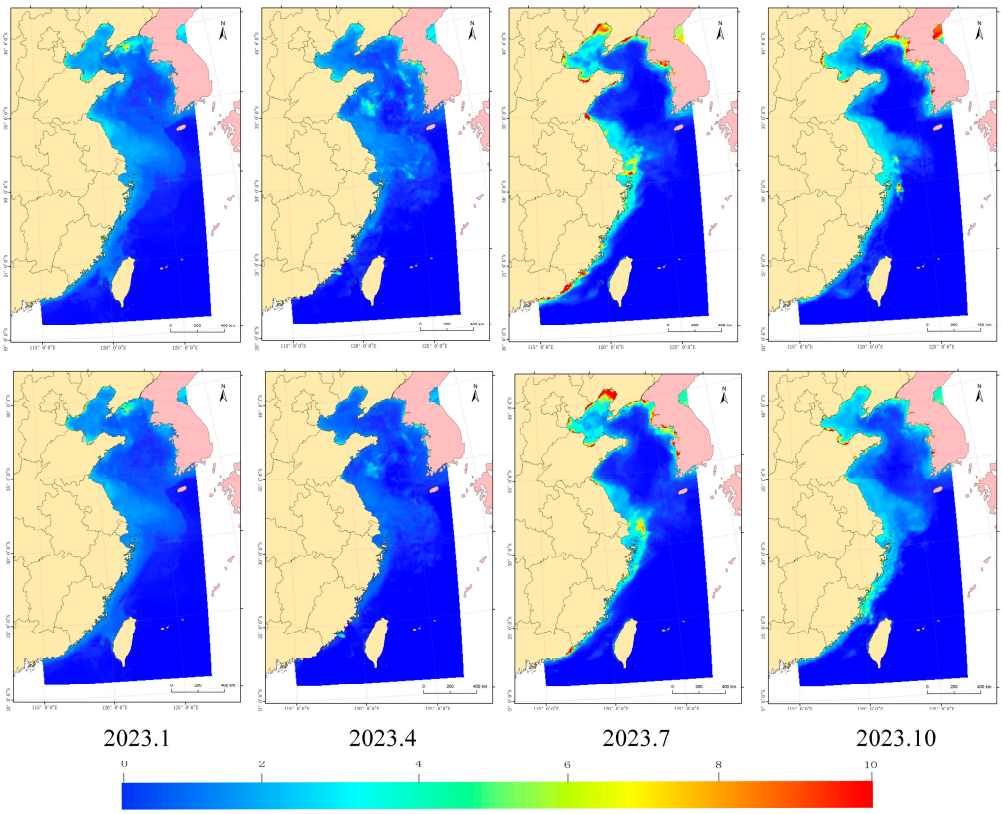


Figure 18. OKG-ConvGRU model long-term prediction results versus observations for time points 2023.1, 2023.4, 2023.7, and 2023.10, with first row of observations and second row of model prediction result images.

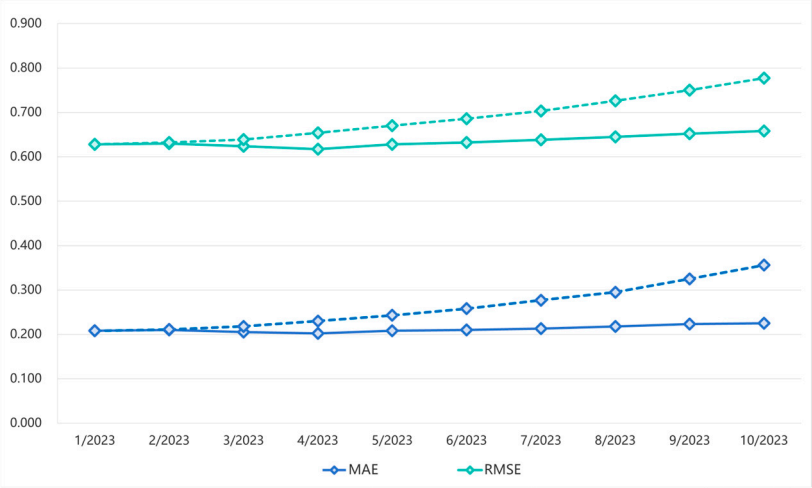


Figure 19. Performance comparison of different long-term prediction methods for the OKG-ConvGRU model. The dashed line represents the use of multi-step rolling forecasts only, while the solid line represents the combination of the Seq2Seq architecture and multi-step rolling forecasts.

4.7. Model Data Efficiency and Robustness Analysis

To explore the impact of the joint knowledge and data-driven approach on data dependency, we compared the prediction performance of the OKG-ConvGRU model with that of the single data-driven CA-ConvGRU model by progressively scaling down the size of the training set. Specifically, the original training set was divided into multiple subsets at different scales (40%, 60%, 80%, and 100%), and both models were trained on each subset and evaluated for their performance on the same test set, as shown in Figure 20. To ensure the reproducibility of the experiments, the same hyperparameter settings were used for all experiments.

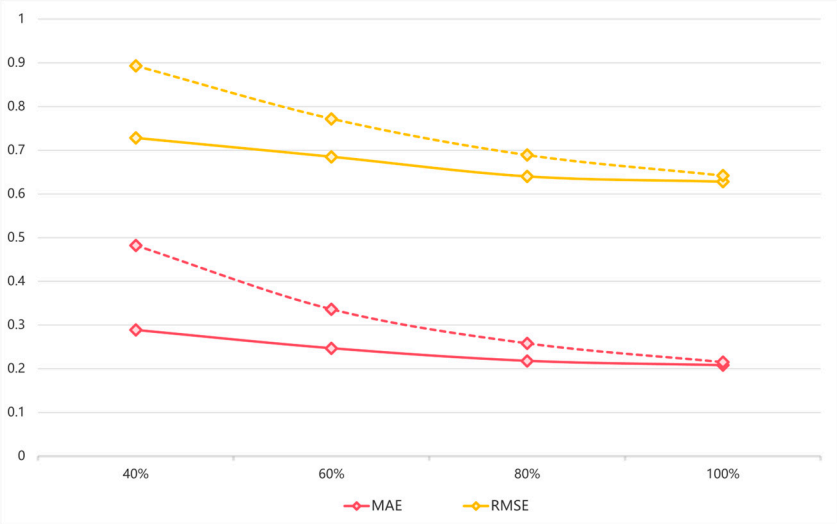


Figure 20. Prediction accuracy versus training data size curves for the OKG-ConvGRU and CA-ConvGRU models. The dashed line represents CA-ConvGRU, while the solid line represents OKG-ConvGRU.

Figure 20 illustrates the dynamic characteristics of the prediction accuracy for both the OKG-ConvGRU and CA-ConvGRU models as functions of training data volume. The experimental results demonstrate that (1) under data reduction condition, the OKG-ConvGRU model exhibits a 28.5% slower increase in MAE compared to CA-ConvGRU, with this advantage becoming more pronounced as data volume decreases, indicating superior data robustness; (2) when trained on the full dataset (100% training data), OKG-ConvGRU achieves a 33.7% lower MAE value than CA-ConvGRU, indicating higher predictive precision; and (3) to attain equivalent prediction performance (MAE $\leq$ 0.25), OKG-ConvGRU requires 24.3% less training data than CA-ConvGRU, significantly enhancing data utilization efficiency. These findings validate the efficacy of integrating knowledge-driven and data-driven strategies in remote sensing spatio-temporal prediction tasks, particularly benefiting scenarios with limited oceanic remote sensing data availability.

5. Conclusions

In this study, we propose a domain knowledge-guided remote sensing prediction framework for ocean elements, which integrates the constructed spatio-temporal knowledge graph of ocean elements and the ConvGRU network. The framework's CAFM employs a cross-attention mechanism to deeply integrate visual and semantic features for ocean elements. This fusion enables the model to leverage both domain knowledge and time-series remote sensing imagery, thereby effectively capturing nonlinear spatial and temporal dependencies among ocean elements. Experimental validation using monthly remote sensing image data of ocean elements in the eastern seas of China (Bohai Sea, Yellow

Sea, and East China Sea) demonstrates the OKG-ConvGRU model's significant advantages over existing benchmarks in prediction accuracy, data utilization efficiency, and long-term prediction stability. The primary research conclusions are as follows:

(1) Joint Knowledge–Data Paradigm: This work introduces the first joint knowledge-and-data-driven remote sensing prediction method for ocean elements, effectively coupling a knowledge graph with a ConvGRU model. The results show that compared to purely data-driven models, this framework not only captures nonlinear spatio-temporal patterns of ocean elements but also elucidates complex inter-element influence mechanisms. By compensating for information gaps that are difficult to infer directly from data, the model achieves marked improvements in prediction accuracy and long-term stability.
(2) Data Efficiency and Robustness: The hybrid knowledge–data approach reduces the model's reliance on large datasets while enhancing its efficiency and robustness. By incorporating knowledge-based constraints and guidance, the model learns accurate patterns from smaller data volumes, mitigating dependency on extensive datasets.
(3) Knowledge Graph Embedding: The semantic representation of the knowledge graph critically influences prediction performance. Specifically, the TransE model demonstrates superior embedding effectiveness compared to TransH for ocean element knowledge graphs, yielding significant overall performance enhancements.
(4) Multi-Step Prediction Strategy: In multi-step forecasting, combining Seq2Seq architecture with multi-step rolling prediction effectively suppresses error accumulation across extended prediction horizons compared to conventional rolling methods.

Although the OKG-ConvGRU framework demonstrates significant performance advantages in marine element prediction, its knowledge embedding mechanism has, to some extent, increased the model complexity and computational overhead. Meanwhile, limited by the size of existing knowledge graphs, the framework currently focuses on spatio-temporal prediction research of typical marine elements (such as chlorophyll-a concentration and sea surface temperature) in the eastern China seas, including the Bohai Sea, East China Sea, and Yellow Sea. Based on the current research status, subsequent work will emphasize the following two aspects: first, optimizing the efficiency of knowledge embedding to reduce the computational load of the model; second, expanding the coverage of the knowledge graph to enhance the framework's practicality and regional adaptability, thereby providing technical support for marine element prediction in more extensive sea areas.

Author Contributions: Conceptualization, Y.C. and R.X.; methodology, R.X. and Y.C.; software, R.X. and D.Z.; validation, R.X.; formal analysis, Y.C. and J.J.; investigation, R.X. and Z.S.; resources, Y.C.; data curation, R.X. and Z.S.; writing—original draft preparation, R.X.; writing—review and editing, R.X. and Y.C.; visualization, R.X. and D.Z.; supervision, Y.C., L.M., and J.J.; project administration, L.M.; funding acquisition, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (415013782), the Natural Science Foundation of Nanjing University of Posts and Telecommunications (NY222173), and the National College Student Innovation Training Program (202410293054Z).

Data Availability Statement: The data supporting the findings of this study are openly available in "figshare" at http://doi.org/10.6084/m9.figshare.28814792 (accessed on 17 April 2025).

Conflicts of 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. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

To clearly show all entities and relationships in OKG, we organized the 146 triples contained in OKG in Table A1.

Table A1. The 146 triplets in OKG.

OKG in Spatial Dimension			OKG in Temporal Dimension		
Head Entity	**Relationship**	**Tail Entity**	**Head Entity**	**Relationship**	**Tail Entity**
Bohai Sea	lat_range	37°23′N–41°23′N	Chl-a	concentration_spring	high
Bohai Sea	lon_range	117°43′E–121°21′E	Chl-a	concentration_summer	low
Bohai Sea	type	semi-enclosed inland sea	Chl-a	concentration_autumn	increase
Bohai Sea	surrounded_by	China	Chl-a	concentration_winter	low
Bohai Sea	north	Liaoning	Chl-a	concentration_march	increase
Bohai Sea	west	Tianjin, Hebei	Chl-a	concentration_april	increase
Bohai Sea	south	Shandong	Chl-a	concentration_july	low
Bohai Sea	connected_to	Yellow Sea via Bohai Strait	Chl-a	concentration_august	low
Bohai Strait	width	76 km	Chl-a	concentration_september	high
Bohai Strait	narrowest_point	Chengshanjiao-Laotieshan	Chl-a	concentration_october	high
Bohai Strait	narrowest_width	34 km	Chl-a	concentration_january	low
Yellow Sea	lat_range	30°20′N–38°43′N	Chl-a	concentration_february	low
Yellow Sea	lon_range	119°30′E–126°18′E	SST	variation_spring	rising
Yellow Sea	type	semi-enclosed marginal sea	SST	variation_summer	highest
Yellow Sea	east_coast	Shandong, Jiangsu	SST	variation_autumn	decreasing
Yellow Sea	west_coast	South Korea, North Korea	SST	variation_winter	lowest
Yellow Sea	connected_to	Bohai Sea via Shandong Peninsula	PIC	variation_spring	rising
Yellow Sea	connected_to	East China Sea via Yangtze Estuary	PIC	variation_summer	decreasing
East China Sea	lat_range	22°42′N–33°33′N	PIC	variation_autumn	increasing
East China Sea	lon_range	118°50′E–127°44′E	PIC	variation_winter	lowest
East China Sea	type	marginal sea	POC	variation_spring	rising
East China Sea	north_coast	Yangtze Estuary, Zhejiang	POC	variation_summer	decreasing
East China Sea	east_coast	Kyushu, Ryukyu Islands	POC	variation_autumn	increasing
East China Sea	south_coast	Fujian, Taiwan	POC	variation_winter	lowest
East China Sea	connected_to	South China Sea via Taiwan Strait	PAR	variation_spring	strengthening
East China Sea	connected_to	Yellow Sea via Yangtze Estuary	PAR	variation_summer	highest

Table A1. *Cont.*

OKG in Spatial Dimension			OKG in Temporal Dimension		
Head Entity	**Relationship**	**Tail Entity**	**Head Entity**	**Relationship**	**Tail Entity**
Yellow River Estuary	lat_lon	37°24′N, 118°52′E	PAR	variation_autumn	weakening
Yellow River Estuary	location	Dongying, Shandong	PAR	variation_winter	lowest
Yellow River Estuary	flows_into	Bohai Sea	NFLH	variation_spring	strengthening
Liao River Estuary	lat_lon	40°38′N, 121°53′E	NFLH	variation_summer	low
Liao River Estuary	location	Panjin, Liaoning	NFLH	variation_autumn	strengthening
Liao River Estuary	flows_into	Bohai Sea	NFLH	variation_winter	lowest
Yalu River Estuary	lat_lon	39°58′N, 124°15′E	kuroshio	strength_spring	strengthening
Yalu River Estuary	location	Dandong, Liaoning-Sinuiju, North Korea	kuroshio	strength_summer	strongest
Yalu River Estuary	flows_into	Yellow Sea	kuroshio	strength_autumn	weakening
Huai River Estuary	lat_lon	33°07′N, 119°53′E	kuroshio	strength_winter	weakest
Huai River Estuary	location	Huai'an, Jiangsu	coastal_current	strength_spring	active
Huai River Estuary	flows_into	Yellow Sea	coastal_current	strength_summer	weakening
Yangtze River Estuary	lat_lon	31°26′N, 121°13′E	coastal_current	strength_autumn	strengthening
Yangtze River Estuary	location	Jiangsu–Shanghai border	coastal_current	strength_winter	stable
Yangtze River Estuary	flows_into	East China Sea	seasonal_ circulation	strength_spring	stable
Qiantang River Estuary	lat_lon	30°12′N, 120°52′E	seasonal_ circulation	strength_summer	seaward
Qiantang River Estuary	location	Hangzhou, Zhejiang	seasonal_ circulation	strength_autumn	northeast
Qiantang River Estuary	flows_into	East China Sea	seasonal_ circulation	strength_winter	shoreward
Min River Estuary	lat_lon	26°00′N, 119°28′E	coastal_current	impact_spring	promote_Chl-a
Min River Estuary	location	Fuzhou, Fujian	coastal_current	impact_autumn	promote_Chl-a
Min River Estuary	flows_into	East China Sea	kuroshio	impact_summer	inhibit_Chl-a
Pearl River Estuary	lat_lon	22°47′N, 113°36′E	seaward_ circulation	impact_summer	inhibit_Chl-a
Pearl River Estuary	location	Guangzhou, Guangdong	shoreward_ circulation	impact_winter	inhibit_Chl-a

Table A1. *Cont.*

OKG in Spatial Dimension			OKG in Temporal Dimension		
Head Entity	**Relationship**	**Tail Entity**	**Head Entity**	**Relationship**	**Tail Entity**
Pearl River Estuary	flows_into	South China Sea	Chl-a	concentration_spring	rising
Chl-a	related_to	SST	Chl-a	concentration_summer	low
Chl-a	higher_in	moderate SST	Chl-a	concentration_autumn	rising
high_low_temp	inhibits	phytoplankton	Chl-a	concentration_winter	low
Chl-a	related_to	PIC			
PIC	positively_ correlated	Chl-a			
high_nutrients	promotes	phytoplankton			
phytoplankton	increases	inorganic_carbon			
Chl-a	related_to	POC			
Chl-a	positively_ correlated	POC			
phytoplankton	affects	organic_carbon			
Chl-a	related_to	PAR			
PAR	drives	phytoplankton_ photosynthesis			
Chl-a	positively_ correlated	PAR			
Chl-a	related_to	NFLH			
NFLH	reflects	Chl-a_fluorescence			
NFLH	positively_ correlated	Chl-a			
high_Chl-a	increases	fluorescence			
Chl-a	higher_in	coastal			
Chl-a	decreases	offshore			
Chl-a	higher_in	enclosed_seas			
Chl-a	trend	Bohai > Yellow Sea > East China Sea			
Yellow Sea	high_Chl-a	unique_water_env			
Bohai Sea	rich_nutrients	land_influence			
Bohai Sea	high_Chl-a	small_size			
East China Sea	uniform_ nutrients	open_water			
East China Sea	high_Chl-a	southern_coast, Taiwan Strait			
Yellow River Estuary	high_Chl-a	rich_nutrients			
Yangtze River Estuary	high_Chl-a	rich_nutrients			
Pearl River Estuary	high_Chl-a	rich_nutrients			

Table A1. *Cont.*

OKG in Spatial Dimension			OKG in Temporal Dimension		
Head Entity	**Relationship**	**Tail Entity**	**Head Entity**	**Relationship**	**Tail Entity**
SST	influenced_by	river_discharge, coastal_currents			
SST	lower_in	coastal			
SST	higher_in	offshore			
SST	trend	coast_to_ offshore_increase			
PIC	higher_in	coastal			
PIC	decreases	offshore			
POC	higher_in	coastal			
POC	decreases	offshore			
PAR	lower_in	coastal			
PAR	higher_in	offshore			
NFLH	stronger_in	coastal			
NFLH	weakens	offshore			
marine_element	active_in	coastal			
marine_element	decrease	offshore			

References

1. Lu, X.; Liu, C.; Niu, Y.; Yu, S.X. Long-Term and Regional Variability of Phytoplankton Biomass and Its Physical Oceanographic Parameters in the Yellow Sea, China. *Estuar. Coast. Shelf Sci.* **2021**, *260*, 107497. [CrossRef]
2. Xing, M.; Yao, F.; Zhang, J.; Meng, X.; Jiang, L.; Bao, Y. Data Reconstruction of Daily MODIS Chlorophyll-a Concentration and Spatiotemporal Variations in the Northwestern Pacific. *Sci. Total Environ.* **2022**, *843*, 156981. [CrossRef] [PubMed]
3. Hu, M.; Ma, R.; Xiong, J.; Wang, M.; Cao, Z.; Xue, K. Eutrophication State in the Eastern China Based on Landsat 35-Year Observations. *Remote Sens. Environ.* **2022**, *277*, 113057. [CrossRef]
4. Sundararaman, H.K.K.; Shanmugam, P. Estimates of the Global Ocean Surface Dissolved Oxygen and Macronutrients from Satellite Data. *Remote Sens. Environ.* **2024**, *311*, 114243. [CrossRef]
5. Gao, H.; Huang, B.; Chen, G.; Li, Y.; Zhang, X.; Wang, L. Deep Learning Solver Unites SDGSAT-1 Observations and Navier–Stokes Theory for Oceanic Vortex Streets. *Remote Sens. Environ.* **2024**, *315*, 114425. [CrossRef]
6. Kim, Y.J.; Kim, H.; Han, D.; Stroeve, J.; Im, J. Long-Term Prediction of Arctic Sea Ice Concentrations Using Deep Learning: Effects of Surface Temperature, Radiation, and Wind Conditions. *Remote Sens. Environ.* **2025**, *318*, 114568. [CrossRef]
7. Silva, R.F.D.; Rijnsdorp, D.P.; Hansen, J.E.; Lowe, R.; Buckley, M.; Zijlema, M. An Efficient Method to Calculate Depth-Integrated, Phase-Averaged Momentum Balances in Non-Hydrostatic Models. *Ocean Model.* **2021**, *165*, 101846. [CrossRef]
8. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep Learning in Environmental Remote Sensing: Achievements and Challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [CrossRef]
9. Xiao, C.; Tong, X.; Li, D.; Chen, X.; Yang, Q.; Xv, X.; Lin, H.; Huang, M. Prediction of Long Lead Monthly Three-Dimensional Ocean Temperature Using Time Series Gridded Argo Data and a Deep Learning Method. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102971. [CrossRef]
10. Baez-Villanueva, O.M.; Zambrano-Bigiarini, M.; Beck, H.E.; McNamara, I.; Ribbe, L.; Nauditt, A.; Birkel, C.; Verbist, K.; Giraldo-Osorio, J.D.; Thinh, N.X. RF-MEP: A Novel Random Forest Method for Merging Gridded Precipitation Products and Ground-Based Measurements. *Remote Sens. Environ.* **2020**, *239*, 111606. [CrossRef]
11. Periasamy, S.; Ravi, K.P.; Tansey, K. Identification of Saline Landscapes from an Integrated SVM Approach from a Novel 3-D Classification Schema Using Sentinel-1 Dual-Polarized SAR Data. *Remote Sens. Environ.* **2022**, *279*, 113144. [CrossRef]
12. Yu, B.; Xu, L.; Peng, J.; Hu, Z.; Wong, A. Global Chlorophyll-a Concentration Estimation from Moderate Resolution Imaging Spectroradiometer Using Convolutional Neural Networks. *J. Appl. Remote Sens.* **2020**, *14*, 034520. [CrossRef]

13. Xie, J.; Zhang, J.; Yu, J.; Xu, L. An Adaptive Scale Sea Surface Temperature Predicting Method Based on Deep Learning with Attention Mechanism. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 740–744. [CrossRef]
14. Skanupong, N.; Xu, Y.; Yu, L.; Wan, Z.; Wang, S. The Convolutional Neural Network for Pacific Decadal Oscillation Forecast. *Environ. Res. Lett.* **2024**, *19*, 124022. [CrossRef]
15. Jia, X.; Ji, Q.; Han, L.; Liu, Y.; Han, G.; Lin, X. Prediction of Sea Surface Temperature in the East China Sea Based on LSTM Neural Network. *Remote Sens.* **2022**, *14*, 3300. [CrossRef]
16. Han, Z.; He, Y.; Liu, G.; Perrie, W. Application of DINCAE to Reconstruct the Gaps in Chlorophyll-a Satellite Observations in the South China Sea and West Philippine Sea. *Remote Sens.* **2020**, *12*, 480. [CrossRef]
17. Wu, Y.; Wang, J.; Zhang, R.; Wang, X.; Yang, Y.; Zhang, T. RIME-CNN-BiLSTM: A Novel Optimized Hybrid Enhanced Model for Significant Wave Height Prediction in the Gulf of Mexico. *Ocean Eng.* **2024**, *312*, 119224. [CrossRef]
18. Farhangi, F.; Sadeghi-Niaraki, A.; Safari Bazargani, J.; Razavi-Termeh, S.V.; Hussain, D.; Choi, S.-M. Time-Series Hourly Sea Surface Temperature Prediction Using Deep Neural Network Models. *J. Mar. Sci. Eng.* **2023**, *11*, 1136. [CrossRef]
19. Zhang, K.; Geng, X.; Yan, X.-H. Prediction of 3-D Ocean Temperature by Multilayer Convolutional LSTM. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1303–1307. [CrossRef]
20. Li, C.; Feng, Y.; Sun, T.; Zhang, X. Long Term Indian Ocean Dipole (IOD) Index Prediction Used Deep Learning by ConvLSTM. *Remote Sens.* **2022**, *14*, 523. [CrossRef]
21. Zhou, G.; Chen, J.; Liu, M.; Ma, L. A Spatiotemporal Attention-Augmented ConvLSTM Model for Ocean Remote Sensing Reflectance Prediction. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *129*, 103815. [CrossRef]
22. 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]
23. Wang, H.; Chen, W.; Li, X.; Liang, Q.; Qin, X.; Li, J. CUG-STCN: A Seabed Topography Classification Framework Based on Knowledge Graph-Guided Vision Mamba Network. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *136*, 104383. [CrossRef]
24. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [CrossRef] [PubMed]
25. Fang, W.; Ma, L.; Love, P.E.D.; Luo, H.; Ding, L.; Zhou, A. Knowledge Graph for Identifying Hazards on Construction Sites: Integrating Computer Vision with Ontology. *Autom. Constr.* **2020**, *119*, 103310. [CrossRef]
26. Zhang, X.; Huang, Y.; Zhang, C.; Ye, P. Geoscience Knowledge Graph (GeoKG): Development, Construction and Challenges. *Trans. GIS* **2022**, *26*, 2480–2494. [CrossRef]
27. Wang, S.; Zhang, X.; Ye, P.; Du, M.; Lu, Y.; Xue, H. Geographic Knowledge Graph (GeoKG): A Formalized Geographic Knowledge Representation. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 184. [CrossRef]
28. Du, J.; Wang, S.; Ye, X.; Sinton, D.S.; Kemp, K. GIS-KG: Building a Large-Scale Hierarchical Knowledge Graph for Geographic Information Science. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 873–897. [CrossRef]
29. Zhu, J.; Dang, P.; Cao, Y.; Lai, J.; Guo, Y.; Wang, P.; Li, W. A Flood Knowledge-Constrained Large Language Model Interactable with GIS: Enhancing Public Risk Perception of Floods. *Int. J. Geogr. Inf. Sci.* **2024**, *38*, 603–625. [CrossRef]
30. Liu, Y.; Ding, J.; Fu, Y.; Li, Y. UrbanKG: An Urban Knowledge Graph System. *ACM Trans. Intell. Syst. Technol.* **2023**, *14*, 60. [CrossRef]
31. Li, Y.; Kong, D.; Zhang, Y.; Tan, Y.; Chen, L. Robust Deep Alignment Network with Remote Sensing Knowledge Graph for Zero-Shot and Generalized Zero-Shot Remote Sensing Image Scene Classification. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 145–158. [CrossRef]
32. Chen, Y.; Dang, X.; Zhu, D.; Huang, Y.; Qin, K. Urban Functional Zone Mapping by Coupling Domain Knowledge Graphs and High-Resolution Satellite Images. *Trans. GIS* **2024**, *28*, 1510–1535. [CrossRef]
33. Tao, Y.; Liu, W.; Chen, J.; Gao, J.; Li, R.; Wang, X.; Zhang, Y.; Ren, J.; Yin, S.; Zhu, X.; et al. A Graph-Based Multimodal Data Fusion Framework for Identifying Urban Functional Zone. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *136*, 104353. [CrossRef]
34. Zhang, Y.; Wang, Y.; Gao, S.; Raubal, M. Context-Aware Knowledge Graph Framework for Traffic Speed Forecasting Using Graph Neural Network. *arXiv* **2024**, arXiv:2407.17703. [CrossRef]
35. Liu, C.; Zhang, X.; Xu, Y.; Xiang, B.; Gan, L.; Shu, Y. Knowledge Graph for Maritime Pollution Regulations Based on Deep Learning Methods. *Ocean Coast. Manag.* **2023**, *242*, 106679. [CrossRef]
36. Liu, X.; Zhang, Y.; Zou, H.; Wang, F.; Cheng, X.; Wu, W.; Liu, X.; Li, Y. Multi-Source Knowledge Graph Reasoning for Ocean Oil Spill Detection from Satellite SAR Images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *116*, 103153. [CrossRef]
37. Geng, Z.; Xu, J.; Wu, R.; Zhao, C.; Wang, J.; Li, Y.; Zhang, C. STGAFormer: Spatial–Temporal Gated Attention Transformer Based Graph Neural Network for Traffic Flow Forecasting. *Inf. Fusion* **2024**, *105*, 102228. [CrossRef]

38. Luo, X.; Song, J.; Guo, J.; Fu, Y.; Wang, L.; Cai, Y. Reconstruction of Chlorophyll-a Satellite Data in Bohai and Yellow Sea Based on DINCAE Method. *Int. J. Remote Sens.* **2022**, *43*, 3336–3358. [CrossRef]
39. Yang, Z.; Chen, J.; Xu, X.; Ran, L.; Jin, H.; Wang, B.; Chen, Q. Seasonal Agricultural Activities and Monsoon Shifts Drive Fluctuations in Nitrogen Levels in Eutrophic Coastal Waters: A Case Study of Xiangshan Bay, China. *Mar. Pollut. Bull.* **2025**, *212*, 117535. [CrossRef]
40. Cen, H.; Jiang, J.; Han, G.; Lin, X.; Liu, Y.; Jia, X.; Ji, Q.; Li, B. Applying Deep Learning in the Prediction of Chlorophyll-a in the East China Sea. *Remote Sens.* **2022**, *14*, 5461. [CrossRef]
41. Zhai, F.; Wu, W.; Gu, Y.; Li, P.; Song, X.; Liu, P.; Liu, Z.; Chen, Y.; He, J. Interannual-Decadal Variation in Satellite-Derived Surface Chlorophyll-a Concentration in the Bohai Sea over the Past 16 Years. *J. Mar. Syst.* **2021**, *215*, 103496. [CrossRef]
42. 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*, 1234567. [CrossRef]
43. Meng, X.; Yao, F.; Zhang, J.; Liu, Q.; Liu, Q.; Shi, L.; Zhang, D. Impact of Dust Deposition on Phytoplankton Biomass in the Northwestern Pacific: A Long-Term Study from 1998 to 2020. *Sci. Total Environ.* **2022**, *813*, 152536. [CrossRef] [PubMed]
44. Chen, Q.; Cai, C.; Chen, Y.; Zhou, X.; Zhang, D.; Peng, Y. TemproNet: A Transformer-Based Deep Learning Model for Seawater Temperature Prediction. *Ocean Eng.* **2024**, *293*, 116651. [CrossRef]
45. Dong, S.; Pavia, F.J.; Subhas, A.V.; Gray, W.R.; Adkins, J.F.; Berelson, W.M. Carbon Cycling in Marine Particles Based on Inorganic and Organic Stable Isotopes. *Geochim. Cosmochim. Acta* **2025**, *388*, 208–220. [CrossRef]
46. Karmakar, J.; Mondal, I.; Hossain, S.A.; Jose, F.; Pichuka, S.; Ghosh, D.; De, T.K.; Lu, Q.-O.; Elkhrachy, I.; Nguyen, N.-M. Analyzing Spatio-Temporal Variability of Aquatic Productive Components in Northern Bay of Bengal Using Advanced Machine Learning Models. Ocean Coast. *Manag.* **2024**, *251*, 107074.
47. McGinty, N.; Guðmundsson, K.; Ágústsdóttir, K.; Marteinsdóttir, G. Environmental and Climactic Effects of Chlorophyll-a Variability around Iceland Using Reconstructed Satellite Data Fields. *J. Mar. Syst.* **2016**, *163*, 31–42. [CrossRef]
48. Wang, T.; Chen, F.; Zhang, S.; Pan, J.; Devlin, A.T.; Ning, H.; Zeng, W. Remote Sensing and Argo Float Observations Reveal Physical Processes Initiating a Winter-Spring Phytoplankton Bloom South of the Kuroshio Current near Shikoku. *Remote Sens.* **2020**, *12*, 4065. [CrossRef]
49. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 2787–2795.
50. Wang, Z.; Zhang, J.; Feng, J.; Chen, Z. Knowledge Graph Embedding by Translating on Hyperplanes. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, Québec City, QC, Canada, 27–31 July 2014.
51. Lin, H.; Cheng, X.; Wu, X.; Yang, F.; Shen, D.; Wang, Z.; Song, Q.; Yuan, W. CAT: Cross Attention in Vision Transformer. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022; pp. 1–6.
52. Wei, X.; Zhang, T.; Li, Y.; Zhang, Y.; Wu, F. Multi-Modality Cross Attention Network for Image and Sentence Matching. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10938–10947.
53. Jian, L.; Xiong, S.; Yan, H.; Niu, X.; Wu, S.; Zhang, D. Rethinking Cross-Attention for Infrared and Visible Image Fusion. *arXiv* **2024**, arXiv:2401.11675. [CrossRef]
54. Helvig, K.; Abeloos, B.; Trouvé-Peloux, P. CAFF-DINO: Multi-Spectral Object Detection Transformers with Cross-Attention Features Fusion. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 17–18 June 2024; pp. 3037–3046.
55. Tian, L.; Li, X.; Ye, Y.; Xie, P.; Li, Y. A Generative Adversarial Gated Recurrent Unit Model for Precipitation Nowcasting. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 601–605. [CrossRef]
56. Kong, W.; Li, H. Remaining Useful Life Prediction of Rolling Bearing under Limited Data Based on Adaptive Time-Series Feature Window and Multi-Step Ahead Strategy. *Appl. Soft Comput.* **2022**, *129*, 109630. [CrossRef]
57. Chen, Y.; Xie, Y.; Dang, X.; Huang, B.; Wu, C.; Jiao, D. Spatiotemporal Prediction of Carbon Emissions Using a Hybrid Deep Learning Model Considering Temporal and Spatial Correlations. *Environ. Model. Softw.* **2024**, *172*, 105937. [CrossRef]
58. Xu, S.; Wang, Y.; Xu, X.; Shi, G.; Zheng, Y.; Huang, H.; Hong, C. A Multi-Step Wind Power Group Forecasting Seq2Seq Architecture with Spatial–Temporal Feature Fusion and Numerical Weather Prediction Correction. *Energy* **2024**, *291*, 130352. [CrossRef]
59. Gao, S.; Zhang, S.; Huang, Y.; Han, J.; Luo, H.; Zhang, Y.; Wang, G. A New Seq2Seq Architecture for Hourly Runoff Prediction Using Historical Rainfall and Runoff as Input. *J. Hydrol.* **2022**, *612*, 128099. [CrossRef]
60. Wang, Y.; Gao, Z.; Liu, D. Multivariate DINEOF Reconstruction for Creating Long-Term Cloud-Free Chlorophyll-a Data Records from SeaWiFS and MODIS: A Case Study in Bohai and Yellow Seas, China. *Remote Sens.* **2019**, *12*, 1383–1395. [CrossRef]

61. Beckers, J.-M.; Barth, A.; Alvera-Azcárate, A. DINEOF Reconstruction of Clouded Images Including Error Maps—Application to the Sea-Surface Temperature around Corsican Island. *Remote Sens. Environ.* **2006**, *102*, 183–199. [CrossRef]
62. Prasetyowati, S.A.D.; Ismail, M.; Budisusila, E.N.; Setiadi, D.R.I.M.; Purnomo, M.H. Dataset Feasibility Analysis Method Based on Enhanced Adaptive LMS Method with Min-Max Normalization and Fuzzy Intuitive Sets. *Int. J. Electr. Comput. Eng.* **2022**, *14*, 55–75. [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.