



# Article Enhancing Dissolved Oxygen Concentrations Prediction in Water Bodies: A Temporal Transformer Approach with Multi-Site Meteorological Data Graph Embedding

Hongqing Wang <sup>1,2</sup>, Lifu Zhang <sup>1</sup>, Rong Wu <sup>3</sup> and Hongying Zhao <sup>4,\*</sup>

- <sup>1</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; wanghongqing21@mails.ucas.ac.cn (H.W.); zhanglf@radi.ac.cn (L.Z.)
- <sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China
- <sup>3</sup> Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China; wr22@mails.tsinghua.edu.cn
- <sup>4</sup> School of Earth and Space Sciences, Peking University, Beijing 100871, China
- Correspondence: zhaohy@pku.edu.cn

Abstract: Water ecosystems are highly sensitive to environmental conditions, including meteorological factors, which influence dissolved oxygen (DO) concentrations, a critical indicator of water quality. However, the complex relationships between multiple meteorological factors from various sites and DO concentrations pose a significant challenge for accurate prediction. This study introduces an innovative framework for enhancing DO concentration predictions in water bodies by integrating multi-station meteorological data. We first construct a dynamic meteorological graph with stationspecific factors as node features and geographic distances as edge weights. This graph is processed using a Geo-Contextual Graph Embedding Module, leveraging a Graph Convolutional Network (GCN) to distill geographical and meteorological features from multi-station data. Extracted features are encoded and then temporally merged with historical DO values to form time-series data. Finally, a Temporal Transformer module is used for future DO concentration predictions. The proposed model shows superior performance compared to traditional methods, successfully capturing the complex relationships between meteorological factors and DO levels. It provides an effective tool for environmental scientists and policymakers in water quality monitoring and management. This study suggests that the integration of graph-based learning and a Temporal Transformer in environmental modeling is a promising direction for future research.

**Keywords:** dissolved oxygen concentration prediction; multi-site meteorological data; graph convolutional networks; temporal transformer; environmental modeling; Tianjin; China

## 1. Introduction

Dissolved oxygen (DO) plays a pivotal role in water environmental science, serving as a critical indicator of the health and sustainability of aquatic ecosystems [1–4]. Oxygen dissolved in water is essential for the survival and growth of aquatic life, including fish, invertebrates, bacteria, and plants [5,6]. Maintaining a healthy balance of DO is essential, as both excessive and inadequate levels can pose severe risks to the ecosystem [7]. High concentrations of DO can lead to the excessive growth of an organism, thereby disrupting the ecosystem balance [8]. Conversely, low DO levels can result in hypoxic conditions that jeopardize aquatic life [9]. Predicting DO concentrations facilitates the management and conservation of aquatic resources, aids in the planning and operation of water treatment processes [10], and helps in the timely detection and mitigation of potential environmental risks [11]. A precise prediction model can offer valuable insights into the future state of the ecosystem, thus providing a powerful tool for decision makers in crafting effective strategies for water resource management and pollution control [12]. Despite its significance, the accurate prediction of DO levels remains a challenging task due to the complexity of aquatic



Citation: Wang, H.; Zhang, L.; Wu, R.; Zhao, H. Enhancing Dissolved Oxygen Concentrations Prediction in Water Bodies: A Temporal Transformer Approach with Multi-Site Meteorological Data Graph Embedding. *Water* **2023**, *15*, 3029. https://doi.org/10.3390/w15173029

Academic Editors: Mehdi Jamei and Masoud Karbasi

Received: 26 July 2023 Revised: 15 August 2023 Accepted: 22 August 2023 Published: 23 August 2023



**Copyright:** © 2023 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/). ecosystems, the multifaceted interactions between numerous influencing factors, and the spatiotemporal variability in DO levels [13].

Existing methods for predicting DO concentrations can generally be classified into three categories: physical models, statistical models, and data-driven models [14]. Physical models are developed based on the physical laws governing DO dynamics, such as the Streeter–Phelps model [15]. These models utilize differential equations to represent the oxygen balance in water, taking into account factors, like biochemical oxygen demand, reaeration, photosynthesis, and respiration [16]. While these models are theoretically sound, they often require a multitude of precise measurements and fail to account for the complex interactions among various environmental factors, making their application, in practice, quite challenging [17,18]. Statistical models, such as regression models, time-series analysis, and Box–Jenkins models, have also been applied to DO prediction [19,20]. These models rely heavily on historical data, and their success depends on the inherent linear relationships among variables [21,22]. However, the interactions between different environmental factors influencing DO levels are complex and often non-linear, which restricts the accuracy and applicability of these models [23–25]. Data-driven models, including machine learning and deep learning models, have gained popularity in recent years owing to their capability to capture complex non-linear relationships and their adaptability to various situations [26–28]. These models, such as artificial neural networks [29,30], support vector machines [31,32], and random forest models [33,34], have shown promising results in DO prediction. However, most existing data-driven models consider only temporal dependencies, overlooking the spatial interactions among different locations, which can lead to suboptimal prediction performance [13,35–37].

To mitigate the limitations of traditional models and incorporate spatial dependencies, recent studies have turned to hybrid models, such as Convolutional Neural Networks (CNNs) combined with Long Short-Term Memory (LSTM) networks [38–40]. CNNs, renowned for their success in image processing tasks, are used to capture spatial correlations by considering the area of interest as an image-like structure [41–43]. Meanwhile, LSTMs handle the temporal dependencies due to their unique architecture that can learn and remember over long sequences [44], alleviating issues encountered with traditional recurrent neural networks, like vanishing or exploding gradients [45].

However, these models present their own set of constraints [46]. Firstly, they usually assume a Euclidean space to capture spatial dependencies [47], which might not accurately reflect the geographical and topological properties of the real-world scenarios, where different meteorological stations and bodies of water exhibit non-Euclidean relationships. Secondly, the heterogeneity of data [48], which is often a combination of structured and unstructured data (like temperature, wind speed, etc.) [49], presents a challenge. Existing models may not fully capture the complex correlations between these different types of data [50]. Moreover, the architecture of these models is rigid, which can hinder the comprehensive incorporation of multiple meteorological factors [51]. These factors, such as temperature, pressure, dew point, wind direction, wind speed, and precipitation, have intricate and non-linear impacts on DO dynamics [52]. For instance, the fixed kernel sizes in CNNs are not conducive to capturing these varying influences and interactions [53], leading to insufficient consideration of environmental factors and, thus, inaccurate predictions.

In light of these limitations, there is a need for a more flexible and sophisticated modeling approach. This approach should accommodate the complex, non-linear interactions among various factors in non-Euclidean spaces and consider both temporal and spatial dependencies simultaneously. This underpins the primary motivation of our study.

The realization of such a sophisticated modeling approach calls for the exploration of advanced data representation and learning techniques. Recently, two techniques have shown exceptional promise in various domains for handling data with complex dependencies and non-Euclidean structures: Graph Embedding techniques [54] and the Transformer model [55]. Graph Embedding techniques have the potential to capture non-Euclidean spatial dependencies effectively [56]. By transforming the nodes of a graph into a continuous vector space while preserving the structural information of the original graph [57], these techniques facilitate the understanding of complex inter-node relationships and topological properties. They have been successfully applied in several domains, including social network analysis [58], bioinformatics [59], and recommendation systems [60], yielding significant improvements over traditional methods. On the other hand, the Transformer model, a deep learning model primarily designed for natural language processing tasks [61], presents a powerful tool for temporal dependency modeling. Its unique self-attention mechanism can effectively capture long-range dependencies in time-series data by weighing the influence of different time steps based on their relevancy [62].

In response to these challenges and opportunities, we propose a novel, sophisticated model for accurate DO concentration predictions that effectively leverages the spatial and temporal dependencies present in multiple meteorological factors across various meteorological stations. Our approach comprises the following: a Meteorological Graph Construction module, wherein meteorological stations are treated as graph nodes; a Geospatial Graph Convolutional Embedding module, applying Graph Convolutional Networks and a Multilayer Perceptron to obtain a comprehensive feature vector; a Feature Encoding and Temporal Concatenation module for feature refining and sequence formation; and, finally, a Temporal Transformer Prediction module, which uses the Transformer's self-attention mechanism for capturing long-term dependencies in data. Our proposed model, tested and validated on real-world data, significantly outperforms existing models, effectively demonstrating its utility in handling the complex, non-linear interactions in DO concentration predictions. We anticipate that this work will contribute to environmental science by improving our understanding of DO dynamics and advancing the predictive modeling techniques in the field.

## 2. Materials

#### 2.1. Study Area

The focus of this investigation is the city of Tianjin, located in the eastern coastal region of North China, approximately 120 km southeast of the capital, Beijing. Encompassing an expansive geographic expanse exceeding 11,300 square kilometers, Tianjin offers a diverse backdrop for studying various meteorological patterns and their impacts on water quality parameters.

Positioned uniquely, Tianjin experiences a humid continental climate influenced by monsoon winds. This climatic influence results in distinct seasonal variations—hot, rainy summers juxtaposed with cold, dry winters. Such seasonality, especially the heavy rainfall during the summer season, significantly influences the region's water bodies. The intricate interrelationship between meteorological factors and the hydrological characteristics of Tianjin's aquatic systems generates a complex environment for the study and prediction of the dissolved oxygen concentrations in these bodies of water. Tianjin boasts a network of diverse water bodies, including rivers, canals, and lakes, interconnected through the prominent Haihe River that empties into the Bohai Sea. Notably, our research pays particular attention to the Binhai New Area, specifically focusing on the Ji Canal Tide Gate within the Haihe River Basin. Given the considerable influence of both natural and human-induced activities on the Ji Canal Tide Gate, it represents a suitable site for examining the typical water quality challenges in the region.

The urbanized nature of Tianjin city, together with its distinct hydrogeological features, introduces multifaceted aspects to its water quality parameters. Our focus, the dissolved oxygen concentrations—a crucial determinant of aquatic health—is impacted by an array of factors. These factors span meteorological elements, including temperature, atmospheric pressure, and wind dynamics, as well as precipitation levels. Moreover, these influencing factors exhibit considerable spatial and temporal variability across Tianjin's vast geography.

Through this investigation, we aim to elucidate the intertwined dynamics between meteorological conditions and water quality, particularly the dissolved oxygen concentrations. This study intends to enhance the predictability of dissolved oxygen levels in Tianjin's water bodies. We anticipate that the insights gleaned from this research will contribute to the academic discourse on water quality prediction and pave the way for more informed, sustainable water resource management practices in Tianjin, aligning with both its ecological imperatives and urban development objectives. Figure 1 illustrates the distribution of the research area and the locations of the data collection sites.



**Figure 1.** Overview of the study area in Tianjin, China. The left panel indicates Tianjin's location within China, while the right panel demonstrates the city's internal administrative divisions. Key features include the Jiyun River in the Haihe River Basin, the focus of our dissolved oxygen concentrations study, and the 11 meteorological stations used for data collection. This illustration underscores the geographic context of our research and the correlation between local meteorological conditions and water quality.

## 2.2. Data Source and Collection

The objectives of this research necessitate the collection of detailed and well-documented datasets from diverse sources. For this study, the data were aggregated from two primary sources: the China National Environmental Monitoring Center (CNEMC) and the United States National Climatic Data Center (NCDC).

The CNEMC, a respected repository of environmental data in China, provides the critical water quality parameter—dissolved oxygen concentrations. The data were systematically harvested from the CNEMC's official online portal (https://szzdjc.cnemc.cn/, accessed on 8 February 2023). The period of data collection extended from 1 January 2021 to 31 December 2022. The data were registered with a temporal resolution of four hours, affording a detailed perspective on the dynamics of dissolved oxygen concentrations over the chosen period. The water quality monitoring station employed for gathering dissolved oxygen concentration data is situated in the Binhai New Area within the Ji Canal Tide Gate of the Haihe River basin, with geographical coordinates of 117.7274° E, 39.1185° N.

On the other hand, meteorological data were sourced from the NCDC (https://www.ncei.noaa.gov/, accessed on 17 April 2023), a preeminent institution within the purview of the National Oceanic and Atmospheric Administration (NOAA), USA. This research incorporated a spectrum of meteorological parameters, including temperature, pressure, dew point, wind direction and speed, and precipitation. Contrasting with the water quality data, meteorological data were captured at a temporal resolution of three hours, thereby yielding a more granular understanding of the atmospheric conditions. The acquisition of meteorological data adhered to the same timeframe as the dissolved oxygen data. Notably,

the data from 2021 served as a foundation for subsequent model training, while the 2022 dataset was reserved for model validation and testing. These datasets, combined, will help in building a robust and accurate predictive model, addressing the complex interplay between water quality and meteorological parameters.

The geographical distribution of the monitoring stations is extensively detailed in Table 1, specifying the monitoring areas, longitude, latitude, sensor altitude, and station altitude for each station.

Category	Station ID	Monitoring Area	Longitude (°E)	Latitude (°N)	Sensor Altitude (m)	Station Altitude (m)
Water Quality Parameter (DO Concentrations)	-	Ji Canal Tide Gate	117.73	39.12	-	-
	54428	Ji County	117.24	40.02	16.9	15.7
	54523	Wuging	117.01	39.23	5.7	4.5
	54525	Baodi	117.17	39.44	6.3	5.1
	54526	Dongli District	117.2	39.05	2.6	1.9
	54527	Tianjin	117.03	39.05	4.3	3.5
Meteorological Factors	54528 54529	Ninghe	117.08 117.49	39.14 39.21	4.6 5.1	3.4 3.9
	54530	Hangu District	117.46	39.14	2.5	1.3
	54622	Jinnan District	117.22	38.59	3.9	3.7
	54623	Tanggu	117.43	39.03	5.7	4.8
	54645	Dagang	117.28	38.51	3.4	2.2

Table 1. Geographical distribution and characteristics of the monitoring stations.

Table 1 offers a detailed view of the distribution of monitoring stations across the research region, with longitude and latitude allowing for precise geographical coordination. The altitudes provide an indication of the vertical profile of the station locations, which may influence certain meteorological and environmental factors. Further, Table 2 provides the specifics of the collected parameters, their respective physical meanings, and their units.

<b>Fable 2.</b> Detailed attributes of the collected parame	ters.
---	-------

Category	Parameter	Physical Meaning	Unit
Water Quality Parameter (DO Concentrations)	Dissolved Oxygen Concentrations	The amount of oxygen dissolved in a unit volume of water	mg/L
Meteorological Factors	Temperature	Degree or intensity of heat present in the substance	°C
	Pressure	The force exerted by the atmosphere at a given point	MPa
	Dew Point	The atmospheric temperature below which water droplets begin to condense and dew can form	°C
	Wind Direction	The direction from which the wind is coming	0
	Wind Speed	The speed at which the air is moving horizontally	m/s
	Precipitation	The amount of rain, snow, or other types of water particles falling from the sky	mm

Table 2 elucidates the physical meanings of the parameters, offering a comprehensive view of their significance in environmental studies. The designated instruments ensure precise data collection, crucial for subsequent data analysis and model development.

#### 2.3. Data Preprocessing

Data preprocessing is a pivotal component in our study, as it ascertains the completeness, consistency, and accuracy of our dataset, thereby ensuring the credibility and robustness of our model predictions.

Our dataset is gathered from two primary sources: meteorological stations and water quality monitoring stations in Tianjin. However, due to various issues, such as equipment failure, station maintenance, and technical complications, the raw dataset might contain missing and outlier values. To counter the potential negative impacts these factors could exert on our analysis results, we implemented a two-step procedure:

- Elimination of missing values: Initially, we discarded records containing missing values that could occur due to issues, like equipment malfunction, station maintenance, or other technical problems. This ensures the completeness of our dataset, enhancing the accuracy of our analysis.
- Removal of outliers: Subsequently, we excluded outlier values from the dataset, i.e., readings significantly deviating from the normal range. Such readings could occur due to equipment malfunction or transient, non-representative environmental conditions. This step aids in reducing data noise and enhances the accuracy of the model's predictions.

Upon cleaning the data, we encountered a crucial issue: aligning data with different time resolutions. Specifically, our meteorological data were recorded every three hours, while water quality parameters (i.e., dissolved oxygen concentrations) were recorded every four hours.

To address this, we adopted a straightforward yet effective method: resampling. We sampled the data every 12 h. This not only solved the inconsistency in time resolution but also made our data more manageable. For the water quality parameters, sampling every 12 h meant that we had two samples each day. Over the span of two years, this would yield approximately 1460 (365 days/year  $\times$  2 years  $\times$  2 samples/day) samples. The meteorological factors were treated similarly, sampled every 12 h.

Through this, we could directly associate the meteorological conditions at each timestamp with the corresponding dissolved oxygen concentrations. Furthermore, through resampling, we preserved the essential information, allowing us to account for temporal variations in meteorological conditions impacting the dissolved oxygen concentrations. This preprocessing step, hence, resolved the inconsistency in time resolution and enriched our data, providing a more comprehensive and detailed input for subsequent analysis and model prediction.

### 3. Methodology

### 3.1. Overview of the Model Architecture

Building predictive models to anticipate changes in critical ecological parameters, such as dissolved oxygen concentrations, is a significant challenge in environmental science and engineering. Addressing this complex task requires processing heterogeneous data sources and accounting for spatial and temporal correlations between these data points.

In this study, we propose a comprehensive model architecture that integrates data from eleven meteorological stations spread across Tianjin City. These stations provide a wealth of meteorological information, including temperature, pressure, dew point, wind direction, wind speed, and precipitation. Our aim is to predict the dissolved oxygen concentrations in the "Production Circle Gate" section of the Tianjin Southern District. Figure 2 presents the entire framework of the model, providing a comprehensive overview of its structure and components.

In our modeling, the subscript notation  $t - t_0 : t - 1$  denotes the time window that starts with  $t - t_0$  and ends at t - 1, encompassing meteorological data from multiple stations and dissolved oxygen concentrations for analysis. Specifically, we denote the time series of meteorological factors for each station  $i \in \{1, 2, ..., n\}$ , where n = 11 corresponds to the number of meteorological stations, as  $X_{t-t_0:t-1}^{(i)}$ . Here, t refers to the current time step, and  $t_0$ , which defaults to 5 in this study, defines the temporal window size. Furthermore,  $D_{t-t_0:t-1}$  represents the historical dissolved oxygen levels at the "Production Circle Gate" section during the same time window. The primary objective of our model is to predict the future dissolved oxygen levels at time t, denoted by  $\hat{Y}_t$ .



**Figure 2.** Architecture of the proposed model for predicting dissolved oxygen concentrations. The figure illustrates the hierarchical structure of our predictive model, starting from the collection of time-series meteorological data from eleven different stations (temperature, pressure, dew point, wind direction, wind speed, precipitation) to the final prediction of dissolved oxygen concentrations. The model includes four primary modules: The Geo-Contextual Graph Embedding Module (which generates a unified feature vector that encapsulates the spatiotemporal characteristics and correlations of the stations), the Feature Encoding and Temporal Concatenation Module (which combines this feature vector with historical dissolved oxygen data at the target site), and the Temporal Transformer Prediction Module (which inputs the combined feature vector into a Transformer model for future dissolved oxygen level prediction). The modular design of the model allows it to capture complex spatiotemporal dynamics in the data effectively.

Our model's design is composed of a layered architecture, systematically processing the data through a sequence of sequential operations represented by the composite function:

$$\hat{Y}_{t} = f\left(h\left(g\left(X_{t-t_{0}:t-1}^{(1)}, X_{t-t_{0}:t-1}^{(2)}, \dots, X_{t-t_{0}:t-1}^{(11)}\right), D_{t-t_{0}:t-1}\right), D_{t-t_{0}:t-1}\right)$$
(1)

This function captures the hierarchical nature of our model concisely. Here,  $g(X_{t-t_0:t-1}^{(1)}, X_{t-t_0:t-1}^{(2)}, \dots, X_{t-t_0:t-1}^{(11)})$  represents the Geo-Contextual Graph Embedding Module. It processes the time-series meteorological factors from the different stations and generates a unified feature vector that encapsulates the spatiotemporal characteristics and correlations of these stations. The function  $h(\cdot, D_{t-t_0:t-1})$  denotes the Feature Encoding and Temporal Concatenation Module. It encodes the feature vector output from the Graph

Embedding Module along with the historical dissolved oxygen data at the target site into a temporally concatenated feature vector. Finally, the function  $f(\cdot, D_{t-t_0:t-1})$  corresponds to the Temporal Transformer Prediction Module. The combined feature vector is input into a Transformer model to predict future dissolved oxygen levels at time *t*.

By structuring our model into these modularized steps, we can account for the complex spatiotemporal dynamics inherent in the meteorological and dissolved oxygen data. The following sections provide a more detailed discussion of the inner workings and motivations behind each module:

- Meteorological Graph Construction Module: This module forms the foundation of our methodology. It capitalizes on the graph-based representation of the meteorological data, enabling us to capture the spatial configuration of meteorological stations and the intricate relations between their respective data.
- Geo-Contextual Graph Embedding Module: As the heart of our model, this module
  provides a sophisticated mechanism for transforming the raw meteorological data into
  a meaningful representation. It utilizes the power of Graph Convolutional Networks
  to process and compress the high-dimensional meteorological data into a lowerdimensional feature vector, capturing both local and global patterns.
- Feature Encoding and Temporal Concatenation Module: This module acts as a bridge between the Geo-Contextual Graph Embedding Module and the Temporal Transformer Prediction Module. It prepares the model's input data by encoding the meteorological feature vector and combining it with the historical dissolved oxygen data, creating a richly informative input for the final prediction module.
- Temporal Transformer Prediction Module: This module is the terminal point of our model architecture. It leverages the potent capabilities of Transformer models in handling sequential data and makes the final prediction of dissolved oxygen concentrations, providing valuable insights for environmental management and policymaking.

Through the intricate combination of graph-based representation, convolutional processing, and transformer-based prediction, our model offers a pioneering approach to predicting dissolved oxygen concentrations using meteorological data. The proposed model is designed to cope with the inherent challenges of environmental data, namely its high dimensionality, complex dependencies, and spatiotemporal variability. With its robust architecture and advanced components, our model stands as a promising tool for environmental monitoring and management.

## 3.2. Meteorological Graph Construction Module

To harness the inherent spatial and temporal correlation between various meteorological stations and effectively feed these into our model, we opt for a Graph Neural Network (GNN)-based representation. GNNs offer a promising approach to capture non-Euclidean characteristics, thereby overcoming the limitations of traditional Convolutional Neural Networks (CNNs) that are primarily designed for Euclidean or grid-like data. Moreover, GNNs are highly capable of embedding heterogeneous data types, which is particularly advantageous given our diverse set of meteorological factors across multiple stations.

Within this framework, each meteorological station is represented as a node in our graph, while the meteorological factors of each station serve as the node's attributes. Let us denote by  $X_{t-t_0:t-1}^{(i)}$  the time series of meteorological factors for each station  $i \in \{1, 2, ..., n\}$ , where n = 11 corresponds to the number of meteorological stations, and  $t - t_0 : t - 1$  denotes the time window under consideration. Then, the attribute tensor  $X_{t-t_0:t-1}$  for our graph can be formed as:

$$\mathbf{X}_{t-t_0:t-1} = \left[ X_{t-t_0:t-1}^{(1)}, X_{t-t_0:t-1}^{(2)}, \dots, X_{t-t_0:t-1}^{(n)} \right]$$
(2)

Furthermore, the connections between these nodes are determined based on the geographic proximity of the stations. Here, we represent these connections using an

adjacency matrix *A*. The matrix *A* is an  $n \times n$  matrix, where  $A_{ij}$  indicates whether there is an edge between node *i* and node *j*.

The geographic proximity is calculated using the Haversine formula, which computes the distance  $d_{ij}$  between two points,  $P_1(lon_1, lat_1)$  and  $P_2(lon_2, lat_2)$ , on the Earth's surface:

$$d_{ij} = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\operatorname{lat}_2 - \operatorname{lat}_1}{2}\right) + \cos(\operatorname{lat}_1)\cos(\operatorname{lat}_2)\sin^2\left(\frac{\operatorname{lon}_2 - \operatorname{lon}_1}{2}\right)}\right)$$
(3)

where *r* is the average radius of the Earth, approximately 6371 km. A threshold of 85 km is then set. If the distance between two stations is less than or equal to this threshold, an edge is created between these two nodes (i.e.,  $A_{ij} = 1$ ). Otherwise, no edge is formed ( $A_{ij} = 0$ ). Formally, the adjacency matrix *A* is defined as follows:

$$A_{ij} = \begin{cases} 1, & \text{if } d_{ij} \le 85 \text{km} \\ 0, & \text{otherwise} \end{cases}$$
(4)

The choice of an 85 km threshold was not made arbitrarily but was determined through extensive experimentation. Figure 3 illustrates the connections between 11 meteorological stations under the optimal threshold condition of 85 km. In Section 4, we present a detailed discussion on the selection of the threshold, highlighting how various options were evaluated to ensure the optimal representation of the spatial correlations between the meteorological stations.



2

**Figure 3.** Visualization of the Meteorological Graph Construction Module. The figure illustrates the network topology of eleven meteorological stations in Tianjin City, each depicted as a node in the graph. Edges are established based on geographic proximity, connecting stations within a 85 km threshold. This diagram exemplifies how the model integrates spatial correlations among meteorological stations into its predictive framework.

With the construction of the adjacency matrix and attribute tensor, our model can efficiently exploit the spatial correlations among the meteorological stations. By leveraging the intrinsic benefits of graph-based representation, the proposed architecture caters to the complexities and nuances of environmental data, offering a sound basis for further stages of the model, including feature encoding and temporal prediction.

## 3.3. Geo-Contextual Graph Embedding Module

Graph embedding methods have garnered significant attention for their prowess in encoding nodes from a graph into a continuous vector space. This facilitates the deeper understanding of not only the graph structure but also the relationships and attributes between nodes. In this research, we specifically harness the power of the Graph Convolutional Network (GCN), a renowned variant of GNN methodologies. The GCN's ability to integrate localized information from the graph structure proves integral in our study.

On the basis of the adjacency matrix and attribute tensor, as developed in Section 3.2, we establish a graph representation *G* consisting of n = 11 nodes. These nodes embody meteorological stations, each possessing a set of specific meteorological attributes. The edges that connect these nodes are determined through geographical proximity and convey the relational network among the stations. The nodes within the graph *G* are symbolized by an attribute tensor  $X_{t-t_0:t}$  with dimensions  $n \times d$ , where *d* represents the dimension of the meteorological factors at each individual station. The adjacency matrix of graph *G*, labelled as *A*, features dimensions  $n \times n$ . In this matrix,  $A_{ij}$  reflects the presence of an edge or connection between stations *i* and *j*.

Within the adopted GCN framework, a graph convolution operation can be mathematically denoted as follows:

$$H^{(l+1)} = \sigma \left( D^{-1} A H^{(l)} W^{(l)} \right)$$
(5)

In this equation,  $H^{(l)}$  signifies the features of nodes at layer l, with  $H^{(0)} = X_{t-t_0:t}$  serving as the initial condition.  $W^{(l)}$  acts as the weight matrix learned during the training phase at layer l. D is the degree matrix with  $D_{ii} = \sum_j A_{ij}$ , and  $\sigma(\cdot)$  acts as a non-linear activation function—in this research, the ReLU function.

The above operation signifies the dual process of feature transformation and neighborhood information aggregation. The newly transformed feature  $H^{(l+1)}$  for each node amalgamates information from its immediate neighbors, providing a localised yet comprehensive summary of the node's context within the graph. By performing this operation over multiple layers, we ensure the assimilation of a more extensive array of contextual information. The output derived from the GCN, denoted by  $H^{(L)}$ , embeds crucial geographical and meteorological contexts for each meteorological station, thereby acting as a potent intermediate representation.

To further hone this intermediate representation, we introduce an additional transformation step facilitated by a Multilayer Perceptron (MLP). The MLP operates as a transformative function  $f_{MLP}(\cdot)$  that maps its input to a higher-dimensional feature space, thereby introducing non-linearity that is capable of capturing complex patterns within the data. The transformation operation of the MLP, which takes  $H^{(L)}$  from the last GCN layer as its input, is as follows:

$$V = \sigma \Big( W^{(2)} \sigma \Big( W^{(1)} H^{(L)} + b^{(1)} \Big) + b^{(2)} \Big)$$
(6)

where,  $W^{(1)}$ ,  $W^{(2)}$  act as weight matrices,  $b^{(1)}$ ,  $b^{(2)}$  denote bias vectors, and  $\sigma(\cdot)$  represents an activation function. This configuration of the MLP facilitates the extraction of higherlevel features from the graph embeddings, transforming the information into a more compact, information-rich representation. The resulting feature vector *V* encapsulates the output of the Geo-Contextual Graph Embedding Module, melding the geographical and meteorological context to provide a robust foundation for subsequent stages of our model.

In combining the GCN and MLP in this module, we effectively capture complex spatiotemporal patterns that are inherent in environmental data, significantly enhancing the representational power of the feature vector *V*. Our Geo-Contextual Graph Embedding Module, through the adept combination of graph-structured data and multilayer neural networks, heralds a novel approach to processing spatiotemporal environmental data.

## 3.4. Feature Encoding and Temporal Concatenation Module

In this section, we delve into the Feature Encoding and Temporal Concatenation Module, a critical stage that bridges the graph-embedded meteorological features and the historical dissolved oxygen concentration measurements, laying the foundation for the subsequent Temporal Transformer Prediction Module. The Geo-Contextual Graph Embedding Module outputs a feature vector *V* for each meteorological station, which is of dimension  $n \times d_v$ .

The first phase of this module entails the positional encoding of the feature vector *V*, employing a one-dimensional Convolutional Neural Network (1D-CNN). The process aims to uncover the latent spatial correlations within the meteorological features. For the feature vector *V*, the 1D-CNN applies a convolution operation, followed by a non-linear activation function; here, we consider the ReLU function. This can be represented mathematically as:

$$P = \sigma(W_{cnn} * V + b_{cnn}) \tag{7}$$

where  $W_{cnn}$  denotes the convolutional kernel,  $b_{cnn}$  is the bias, \* represents the convolution operation, and  $\sigma(\cdot)$  is the ReLU activation function. The output *P* is the positionally encoded feature vector of dimension  $n \times d_p$ , where  $d_p$  is the dimension of the feature vector after CNN encoding.

The next step integrates the positionally encoded feature vectors with historical measurements of the dissolved oxygen concentrations. Let  $D_{t-t_0:t-1}$  be the historical dissolved oxygen concentrations from time  $t - t_0$  to t - 1, which is of dimensions  $1 \times t_0$ . These historical measurements are appended to the positionally encoded feature vector P, forming a temporally concatenated matrix T of dimensions  $n \times (d_p + t_0)$ :

$$Z = \begin{bmatrix} P, D_{t-t_0:t-1} \end{bmatrix} \tag{8}$$

where the brackets denote the concatenation operation, and the output *Z* represents the temporally concatenated matrix.

The temporally concatenated matrix Z is a comprehensive time-series dataset that fuses the positionally encoded meteorological features with historical dissolved oxygen concentrations. The data are now prepared for the next module, the Temporal Transformer Prediction Module, that predicts the future dissolved oxygen concentrations. This Feature Encoding and Temporal Concatenation Module effectively brings together the meteorological data and dissolved oxygen history and instills a temporal facet to the model.

## 3.5. Temporal Transformer Prediction Module

The Temporal Transformer Prediction Module, drawing on the power of the Transformer model [55], forms the heart of our proposed system, synthesizing the preceding stages' outputs to predict future dissolved oxygen concentrations. The Transformer model's pivotal strength, the self-attention mechanism, demonstrates proficiency in modeling both local and long-range dependencies in sequences, rendering it aptly suitable for our predictive task from the spatio-temporal environmental data. The Transformer architecture can be conceptually segregated into two primary components, the Encoder and Decoder, both of which are formed of multiple identical layers, each consisting of a self-attention mechanism and a position-wise feed-forward network.

The Encoder operates as a complex, sequential data interpreter. The input to the Encoder is the combined feature sequence *Z*, and its output is a high-dimensional representation of the input. This output is a result of the input sequence flowing through layers of self-attention mechanisms and feed-forward networks.

A critical aspect of the self-attention mechanism within the Encoder is the implementation of a mask over the future time steps. This mask is applied to prevent the attention mechanism from incorporating information about future dissolved oxygen concentrations during the training process. This ensures that the predictions made by the model are based solely on past and current information, thereby preserving the temporal sequence's integrity and avoiding any data leakage that would artificially enhance the model's performance.

Following the Encoder, the Decoder takes the encoded sequence and uses it alongside its own prior outputs to generate the future sequence for dissolved oxygen concentration levels. The Decoder has its own self-attention mechanism that allows it to recognize patterns in the sequence it is generating while simultaneously considering the encoded information. This dual mechanism enhances the prediction process by enabling the Decoder to be aware of the broader context, thereby improving the overall predicting accuracy.

The self-attention mechanism can be formulated as:

Attention
$$(Q, K, V) = \operatorname{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$
 (9)

where Q, K, and V represent the query, key, and value matrices, respectively, and  $d_k$  is the dimensionality of the keys. The mechanism calculates attention scores based on the compatibility of each query with each key. These scores are then used to form a weighted sum of the values.

The final step of the Temporal Transformer Prediction Module is the application of a linear transformation layer on the output from the Decoder. This transformation produces the corresponding prediction for the future dissolved oxygen concentrations:

$$DO_{pred} = W_o \cdot \mathbf{D} + b_o \tag{10}$$

where  $W_o$  and  $b_o$  represent the weight matrix and the bias term of the final linear layer, respectively. The length of the predicted sequence  $DO_{pred}$  equals the predicting step of the dissolved oxygen concentrations, which forms the final output of the Temporal Transformer Prediction Module.

In summary, the Temporal Transformer Prediction Module effectively captures and utilizes local and global spatio-temporal dependencies in the input data to provide accurate predictions of future dissolved oxygen concentrations. This contributes to effective water quality prediction and holds substantial potential in aiding environmental science research and water quality management.

#### 3.6. Model Configuration and Experimental Framework

In the preceding subsections, we detailed the construction of our novel environmental data prediction model, which comprises four main modules: the Meteorological Graph Construction module, the Geo-Contextual Graph Embedding module, the Feature Encoding and Temporal Concatenation module, and the Temporal Transformer Prediction module. Herein, we designate this model as the Meteorological Graph and Temporal Transformer, abbreviated as MegaTT, to accurately describe its primary features and functionality. Table 3 provides an overview of the key parameter settings for each module in the MegaTT model.

**Table 3.** Overview of key parameter settings in the MegaTT model.

Module	Parameter	Setting
Meteorological Graph Construction Module	Number of Neighbors	5
Geo-Contextual Graph Embedding Module	Embedding Size	64
Feature Encoding and Temporal	Encoded Feature Size	128
Concatenation Module	Temporal Window Size	5
	Number of Attention Heads	4
Temporal Transformer	Size of Hidden States	256
Prediction Module	Number of Encoder Layers	2
	Number of Decoder Layers	2

In addition to the MegaTT model's architectural parameters, the optimization strategies utilized during the training process also play a significant role in determining the predictive performance. This section delineates the particular parameters associated with the training procedure, ranging from the selected loss functions to optimizers. The dataset, which was employed for model training and testing, spans from 1 January 2021 to 31 December 2022, furnishing a total of 1460 data samples with the assumption of two collected samples per day. The chronological segmentation of the data into training and testing sets involved utilizing the first year's data for training and the subsequent year's data for model validation.

A sliding window approach was incorporated during the model's training, wherein each window consisted of five sequential data samples (equivalent to a 2.5-day duration) employed to forecast the following data sample. This approach determined the prediction window and step length to be equivalent to one data sample or a half-day duration, indicating that the previous 2.5 days of meteorological data were utilized to forecast the weather conditions for the following half-day duration.

The selection of the optimizer, loss function, and other related training parameters is crucial for securing optimal model performance. These parameters were ascertained based on a systematic series of optimization trials. Table 4 displays some of the settings used during the model training, providing details on the specific configurations and parameters.

Table 4. Configuration specifics utilized for training the MegaTT model.

Training Parameter	Value
Optimizer	Adam
Loss Function	Mean Squared Error
Learning Rate	0.001
Batch Size	64
Training Parameter	Value

In the wake of the aforementioned training and optimization, we employ several key metrics to evaluate the predictive performance of our model, MegaTT. These include the Root Mean Square Error (*RMSE*), Mean Absolute Error (*MAE*), and the coefficient of determination,  $R^2$ . Both *RMSE* and *MAE* provide us with measures of prediction error, while  $R^2$  provides a measure of the explanatory power of the model. Their equations are as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$
(11)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$
(12)

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - \hat{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}$$
(13)

where  $y_i$  is the actual value,  $\hat{y}_i$  is the predicted value,  $\overline{y}$  is the average of the actual values, and *n* is the total number of observations.

In the upcoming Section 4.1, we will conduct a comparative study between our MegaTT model and other extant models for dissolved oxygen concentration predictions. This comparison will focus on contrasting the predictive performance, thereby underscoring the superior attributes of the MegaTT model. In Section 4.2, we delve into the discussion on how variations in the distance threshold in meteorological graph construction impact model performance. Section 4.3 outlines an ablation study of the meteorological module, assessing the spatial impact of meteorological features on by using only the Temporal Transformer for dissolved oxygen concentration predictions. Additionally, we investigate the model's behavior using the nearest single-station approach, focusing on the impact of the nearest meteorological station within the deep

learning method. This exploration will help us better understand the key drivers of model performance and potential avenues for optimization.

## 4. Results and Discussion

## 4.1. Performance Analysis and Model Comparison

In this section, we provide a detailed comparison of our proposed model, MegaTT, with a series of established benchmark models extensively utilized in the field of environmental data prediction. For the sake of transparency, reproducibility, and fairness in our experimental setup, each model's specific architectural configurations and principal parameters are comprehensively explained.

- Support Vector Machine (SVM) [32]: The SVM model implemented in this study utilized the Radial Basis Function (RBF) kernel to map the input space into a higher dimension. The penalty parameter C and kernel coefficient gamma, both essential to the SVM's operation, were finetuned through a grid search in the range of {0.1, 1, 10, 100}, with the goal of minimizing the prediction error on a separate validation set.
- Random Forest (RF) [33]: The RF model, a robust ensemble learning method, was employed with varying numbers of decision trees. The optimal number of trees, chosen from the set {100, 200, 500, 1000}, was determined via cross-validation to mitigate overfitting and to ensure that the model generalized well to unseen data.
- Extreme Gradient Boosting (XGBoost) [37]: The XGBoost model, renowned for its predictive power and efficiency, was configured with a learning rate of 0.1, a maximum tree depth of 5, and 100 estimators. Further finetuning of these parameters was performed based on a validation set to optimize the balance between learning speed and prediction accuracy.
- Long Short-Term Memory (LSTM) [34]: The LSTM model was implemented with a twolayer architecture, each layer comprising 50 units, to capture temporal dependencies. A dropout rate of 0.2 was introduced to control overfitting, thus preventing the model from excessively relying on particular features or training instances.
- Gated Recurrent Unit (GRU) [21]: The GRU model, a variant of the recurrent neural network, was utilized with a single hidden layer composed of 100 units. Similar to LSTM, a dropout rate of 0.2 was applied to maintain model generalization.

In order to ensure a fair and unbiased comparison, all models were trained and evaluated using an identical dataset. The determination of hyperparameters was guided by grid search, coupled with cross-validation on the training data. The performance of the models was gauged using the same metrics *RMSE*, *MAE*, and *R*<sup>2</sup>, providing a holistic and comprehensive evaluation of their prediction accuracy and generalizability.

Figure 4 provides a visual time-series analysis, highlighting the comparative predictive accuracy of MegaTT and the benchmark models. It can be observed that the MegaTT model captures the variations in dissolved oxygen concentrations effectively, closely following the observed values. In comparison, although all models demonstrate competence in capturing the general trend, they fail to capture sudden changes or maintain consistent accuracy across the time span, a challenge efficiently tackled by our proposed MegaTT model.

The comparative error distribution of all models is illustrated in Figure 5. Here, we can observe that MegaTT's prediction errors are mostly concentrated in lower error intervals, implying higher prediction accuracy. In contrast, other models demonstrate wider error distributions, indicating less stable predictive performances.

The comparative performance metrics are further tabulated in Table 5. These results provide quantitative evidence to the superiority of MegaTT over the benchmark models. Specifically, MegaTT outperforms all other models with the lowest *RMSE* and *MAE* and the highest  $R^2$  score. This implies that our MegaTT model achieves superior precision, lower tendency of large errors, and better consistency with the observed data.



**Figure 4.** A Comparative time-series analysis of predicted and observed dissolved oxygen concentrations for 2022. Each sub-figure demonstrates the predictive accuracy of a specific model: (**a**) Meteorological Graph and Temporal Transformer (MegaTT), a novel approach utilizing graph embeddings of meteorological stations coupled with a temporal transformer for optimized predictions; (**b**) Support Vector Machine (SVM); (**c**) Random Forest (RF); (**d**) Extreme Gradient Boosting (XGBoost); (**e**) Long Short-Term Memory (LSTM); (**f**) Gated Recurrent Unit (GRU). The composition of these individual models in one figure allows for a rigorous and direct comparison of their predictive performance.



**Figure 5.** Histogram of prediction errors for different models. The x-axis represents different ranges of prediction errors while the y-axis shows the frequency within each range. Each range of errors includes six bars, each corresponding to a different model: Meteorological Graph and Temporal Transformer (MegaTT), Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The bars' heights in each range indicate the frequency of errors falling within that range for the respective model. The midpoints of the tops of the MegaTT bars across all error ranges are connected by a line, providing a visual representation of the error distribution specific to the MegaTT model.

Model	RMSE	MAE	$R^2$
MegaTT	0.754	0.601	0.936
SVM	1.502	1.180	0.749
RF	1.634	1.238	0.711
XGBoost	1.480	1.137	0.756
LSTM	1.484	1.113	0.753
GRU	1.411	1.069	0.775

Table 5. Comparative performance metrics of MegaTT and other models.

In conclusion, through the rigorous comparisons in this section, it is unequivocally demonstrated that our proposed MegaTT model outperforms a variety of established models in the field. The superior performance is evident in both the time-series analysis and the error distribution, and it is further corroborated by the performance metrics. This suggests that our novel approach of employing the graph embeddings of meteorological stations, coupled with a temporal transformer for optimized predictions, provides a more robust and accurate tool for predicting dissolved oxygen concentrations in water bodies.

## 4.2. Impact of Meteorological Graph Connectivity Variation on Model Performance

The role of meteorological graph connectivity in enhancing the prediction performance of the MegaTT model is explored in this section. In an effort to understand the effects of varying graph connectivity on the model's performance, we incrementally adjusted the spatial threshold from zero (signifying no connections among vertices) up to 170 km (a distance surpassing the longest inter-station gap, thereby resulting in a fully connected graph).

Figure 6 serves as a heatmap of the spatial distances between the 11 meteorological stations and the resultant impact on model performance is illustrated in Figure 7 and Table 6. Figure 7's boxplot indicates the absolute prediction errors of the model corresponding to each spatial threshold. Each box's median value is linked to visualize the trend of error variation with the changing threshold.

Threshold (km)	RMSE	MAE	$R^2$
$\mathbf{r} = 0$	1.333	1.069	0.824
r = 17	1.345	1.078	0.822
r = 34	0.967	0.742	0.899
r = 51	0.841	0.676	0.919
r = 68	0.790	0.624	0.931
r = 85	0.754	0.601	0.936
r = 102	0.809	0.628	0.927
r = 119	0.793	0.627	0.928
r = 136	0.844	0.663	0.922
r = 153	1.212	0.952	0.849
r = 170	1.196	0.948	0.845

Table 6. Performance metrics of the model under different spatial thresholds.

Notably, as the spatial threshold increased from 0 to 85 km, the model's Root Mean Squared Error (*RMSE*) and Mean Absolute Error (*MAE*) exhibited a general downward trend, despite slight fluctuations. Concurrently, the coefficient of determination ( $R^2$ ) increased. This outcome suggests that, within this range, expanding the spatial threshold—thus broadening the information exchange between meteorological stations—enhances the model's predictive accuracy for dissolved oxygen concentrations. It could be due to the fact that meteorological stations within this range likely capture similar atmospheric conditions, which contribute significantly to the dissolved oxygen concentration levels in the associated water bodies.

							0-0-0							
54428	- 0.00	90.03	64.77	107.91	109.35	98.81	92.58	99.65	159.02	111.28	167.94		- 160	
54523	- 90.03	0.00	27.10	25.87	20.09	11.69	41.41	40.05	73.45	42.51	83.40		- 140	
54525	- 64.77	27.10	0.00	43.44	45.01	34.25	37.57	41.66	94.61	50.79	103.85			
54526 □	- 107.91	25.87	43.44	0.00	14.68	14.40	30.70	24.57	51.18	19.99	60.44		- 120	
I 54527	- 109.35	20.09	45.01	14.68	0.00	10.90	43.48	38.43	53.73	34.62	63.84	-	- 100	n)
S 54528	- 98.81	11.69	34.25	14.40	10.90	0.00	36.19	32.77	62.35	32.59	72.16		- 80	stance (ki
eteorol 54529	- 92.58	41.41	37.57	30.70	43.48	36.19	0.00	8.20	72.79	20.67	79.93		ć	Dig
≥ 54530	- 99.65	40.05	41.66	24.57	38.43	32.77	8.20	0.00	64.59	12.50	71.77		- 60	
54622	- 159.02	73.45	94.61	51.18	53.73	62.35	72.79	64.59	0.00	52.20	10.31		- 40	
54623	- 111.28	42.51	50.79	19.99	34.62	32.59	20.67	12.50	52.20	0.00	59.27	-	- 20	
54645	- 167.94	83.40	103.85	60.44	63.84	72.16	79.93	71.77	10.31	59.27	0.00			
	54428	54523	54525	54526	54527	54528	54529	54530	54622	54623	54645		- 0	

Distances between Meteorological Stations

Meteorological Station ID

**Figure 6.** Heatmap of distances between meteorological stations. This figure presents a distance matrix as a heatmap, where each cell represents the distance in kilometers between pairs of meteorological stations. The station pairs are identified by their unique Station ID. Colors in the heatmap range from light (representing shorter distances) to dark (representing longer distances). The diagonal line from the top-left to the bottom-right, which shows the self-distance of each station, is colorless as it represents a distance of zero.

However, when the spatial threshold extended beyond 85 km, the *RMSE* and *MAE* began to climb while the  $R^2$  declined. This suggests that an excessively high threshold, leading to a fully connected graph, may introduce irrelevant associations. For example, meteorological stations that are geographically distant might not share similar atmospheric conditions, and forcibly connecting them could introduce noise into the model. Consequently, this unnecessary information exchange may distort the model's learning, causing a decline in prediction performance.

Table 6 corroborates the observations from Figure 7, revealing an optimal threshold of 85 km for constructing the meteorological graph, a key factor in the MegaTT model's performance. This optimal threshold is vital for capturing the spatial-temporal structure of multi-site meteorological data. When the threshold was below 85 km, suburban meteorological stations 54622, 54645, and 54428 gradually disconnected from the main central city area meteorological graph, impacting the prediction accuracy. Conversely, when the threshold

exceeded 85 km, the connections became overly dense, introducing noise and distorting the model's learning. An 85 km threshold allowed for connections that were neither too dense nor too sparse, achieving the highest prediction accuracy for dissolved oxygen concentrations in the target area, located in the central city. This finding underscores the importance of a well-defined threshold for the meteorological graph, emphasizing the significance of balanced connections between remote suburban and central city meteorological stations. It not only enhances the prediction of dissolved oxygen concentrations but also provides essential guidance for future research on graph-structured environmental data analysis, laying a foundation for a more precise and robust modeling of environmental phenomena.



**Figure 7.** Impact of varying spatial thresholds on model performance. The boxplots represent the distribution of absolute prediction errors of the MegaTT model at different spatial thresholds. The line linking the medians of each boxplot visualizes the trend of error variation with the changing threshold. The figure illustrates how the model's performance varies with spatial thresholds, highlighting the existence of an optimal threshold that maximizes predictive accuracy.

#### 4.3. Ablation Study of Meteorological Module and Impact of Nearest Single-Station Approach

In this section, two principal experiments are designed to evaluate the effectiveness of the proposed MegaTT model in capturing meteorological factors that influence dissolved oxygen concentration predictions in water bodies.

- The first experiment involves a comparison between the full MegaTT model and its reduced form, referred to as the Temporal Transformer (TT) model. By eliminating all meteorological station inputs and retaining only the Temporal Transformer Prediction Module, this comparison serves to highlight the contributions of the integrated meteorological modules in the MegaTT model.
- The second experiment emphasizes a specific configuration, where only the meteorological factors from the nearest station are retained in the MegaTT model (NS-MegaTT). Table 7 lists the distances between 11 meteorological stations and the target water quality monitoring site, highlighting the proximate relationships. Among them, station 54529 is identified as the closest to the target water quality monitoring site. By employing data exclusively from this nearest meteorological station, the analysis aims to assess the impact of nearest single-station information on the model's capability in predicting dissolved oxygen concentrations accurately.

Station ID	Monitoring Area	Distance (km)
54428	Ji County	108.599
54523	Wuqing	63.071
54525	Baodi	59.831
54526	Dongli District	46.154
54527	Tianjin	60.674
54528	Beichen District	55.894
54529	Ninghe	22.857
54530	Hangu District	23.189
54622	Jinnan District	73.375
54623	Tanggu	27.494
54645	Dagang	77.979

**Table 7.** Distances between the water quality monitoring station located at Ji Canal Tide Gate and 11 meteorological stations.

Figure 8 provides a visual representation of the absolute errors for each model. The grey dashed lines marking the *MAE* in each subplot further elucidate the differences between the models.



**Figure 8.** Time-series plots of the absolute errors between predicted and actual dissolved oxygen concentrations for each sample point in the test dataset, depicted for three distinct models: Temporal Transformer (TT), Nearest Single-Station MegaTT (NS-MegaTT), and full MegaTT. Each subplot presents the error over time, with a grey dashed line marking the Mean Absolute Error (*MAE*) for the corresponding model. The comparison illustrates the efficiency and accuracy of the models in predicting dissolved oxygen concentrations, highlighting the contribution of the meteorological modules in the MegaTT model.

The TT model, without meteorological inputs, displays higher errors (*RMSE*: 1.427, *MAE*: 1.154) and a relatively lower determination coefficient ( $R^2$ : 0.798). NS-MegaTT, which incorporates data only from the nearest meteorological station, improves performance slightly (*RMSE*: 1.236, *MAE*: 0.973,  $R^2$ : 0.848). The full MegaTT model shows substantial improvements with the lowest errors (*RMSE*: 0.754, *MAE*: 0.601) and the highest determination coefficient ( $R^2$ : 0.936). The comparison emphasizes the contribution of multistation meteorological data in predicting dissolved oxygen concentrations. The MegaTT model's superiority is apparent, with significantly better performance in terms of error minimization and  $R^2$  value.

The results of the ablation study offer profound insights into the importance of meteorological data integration. The MegaTT model's significant enhancement in prediction accuracy underscores the effectiveness of the dynamic meteorological graph construction.

#### 5. Conclusions

This study aimed to enhance the prediction of dissolved oxygen concentrations in water bodies through a novel approach, the Meteorological Graph and Temporal Transformer (MegaTT) model. This model effectively exploited the spatial-temporal structure of multi-site meteorological data, providing a comprehensive understanding of the water bodies' characteristics and their impacts on dissolved oxygen concentration predictions.

Our MegaTT model outperformed traditional machine learning models, including Support Vector Machine (SVM), Random Forests (RFs), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), demonstrating its superiority in handling complex geospatial and temporal patterns.

Furthermore, the study determined an optimal distance threshold of 85 km for constructing the meteorological graph, achieving the highest prediction accuracy for dissolved oxygen concentrations in the target area. This threshold maintained a balanced connection between remote suburban and central city meteorological stations. This finding emphasizes the importance of carefully defining connections within the graph and provides essential guidance for future research on graph-structured environmental data analysis.

An ablation study further underscored the essential role of the meteorological module, with the MegaTT model significantly outperforming its reduced versions, namely the Temporal Transformer (TT) model and the Nearest Single-Station MegaTT (NS-MegaTT). This in-depth analysis provided a robust justification for integrating multi-station meteorological data, leading to improved dissolved oxygen prediction accuracy.

The MegaTT model presented in this paper opens up a new perspective on dissolved oxygen concentration predictions. It not only shows promising results but also paves the way for the potential incorporation of other environmental factors, advancing the development of holistic and effective water quality management strategies. Future research directions could involve exploring other types of environmental data and applying the MegaTT model to different water quality parameters, which would be of great significance for environmental management and policymaking.

**Author Contributions:** Conceptualization, H.W. and L.Z.; methodology, H.W.; software, H.W.; validation, H.W., L.Z., R.W. and H.Z.; formal analysis, H.W.; investigation, H.W. and H.Z.; resources, H.W. and H.Z.; data curation, H.W. and H.Z.; writing—original draft preparation, H.W.; writing—review and editing, H.W., L.Z., R.W. and H.Z.; visualization, H.W.; supervision, H.W. and H.Z.; project administration, H.W.; funding acquisition, L.Z. 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, grant number 41830108.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

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

## References

- Zuo, Y.; Chen, L.; Hu, X.; Wang, F.; Yang, Y. Silver Nanoprism Enhanced Colorimetry for Precise Detection of Dissolved Oxygen. *Micromachines* 2020, 11, 383. [CrossRef]
- Khullar, S.; Singh, N. Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation. Environ. Sci. Pollut. Res. 2022, 29, 12875–12889. [CrossRef] [PubMed]
- Liu, P.; Wang, J.; Sangaiah, A.K.; Xie, Y.; Yin, X. Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. *Sustainability* 2019, 11, 2058. [CrossRef]
- 4. Solarski, M.; Rzętała, M. Ice Regime of the Kozłowa Góra Reservoir (Southern Poland) as an Indicator of Changes of the Thermal Conditions of Ambient Air. *Water* **2020**, *12*, 2435. [CrossRef]
- 5. Nakova, E.; Linnebank, F.E.; Bredeweg, B.; Salles, P.; Uzunov, Y. The river Mesta case study: A qualitative model of dissolved oxygen in aquatic ecosystems. *Ecol. Inform.* **2009**, *4*, 339–357. [CrossRef]
- 6. Solarski, M.; Pradela, A.; Rzetala, M. Oxygen conditions in anthropogenic lakes of the silesian upland (southern poland). *Int. Multidiscip. Sci. GeoConf. SGEM* **2012**, *3*, 785.
- Baxa, M.; Musil, M.; Kummel, M.; Hanzlík, P.; Tesařová, B.; Pechar, L. Dissolved oxygen deficits in a shallow eutrophic aquatic ecosystem (fishpond)—Sediment oxygen demand and water column respiration alternately drive the oxygen regime. *Sci. Total Environ.* 2021, 766, 142647. [CrossRef]
- Kita, Y.; Nishikawa, H.; Takemoto, T. Effects of cyanide and dissolved oxygen concentration on biological Au recovery. J. Biotechnol. 2006, 124, 545–551. [CrossRef]
- 9. Kramer, D.L. Dissolved oxygen and fish behavior. Environ. Biol. Fishes 1987, 18, 81–92. [CrossRef]
- 10. Li, D.; Zou, M.; Jiang, L. Dissolved oxygen control strategies for water treatment: A review. Water Sci. Technol. 2022, 86, 1444–1466. [CrossRef]
- 11. Kisi, O.; Alizamir, M.; Docheshmeh Gorgij, A. Dissolved oxygen prediction using a new ensemble method. *Environ. Sci. Pollut. Res.* **2020**, *27*, 9589–9603. [CrossRef]
- 12. Maier, H.R.; Jain, A.; Dandy, G.C.; Sudheer, K.P. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* **2010**, *25*, 891–909. [CrossRef]
- 13. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
- 14. Zhi, W.; Feng, D.; Tsai, W.-P.; Sterle, G.; Harpold, A.; Shen, C.; Li, L. From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale? *Environ. Sci. Technol.* 2021, 55, 2357–2368. [CrossRef] [PubMed]
- 15. Rinaldi, S.; Soncini-Sessa, R.; Romano, P. Parameter Estimation of Streeter-Phelps Models. J. Environ. Eng. Div. 1979, 105, 75–88. [CrossRef]
- 16. Gotovtsev, A.V. Modification of the Streeter-Phelps system with the aim to account for the feedback between dissolved oxygen concentration and organic matter oxidation rate. *Water Resour.* **2010**, *37*, 245–251. [CrossRef]
- 17. Nas, S.S.; Nas, E. Water Quality Modeling and Dissolved Oxygen Balance in Streams: A Point Source Streeter-Phelps Application in the Case of the Harsit Stream. *CLEAN—Soil Air Water* **2009**, *37*, 67–74. [CrossRef]
- 18. Wu, J.; Yu, X. Numerical Investigation of Dissolved Oxygen Transportation through a Coupled SWE and Streeter–Phelps Model. *Math. Probl. Eng.* **2021**, 2021, 6663696. [CrossRef]
- 19. Huck Peter, M.; Farquhar Grahame, J. Water Quality Models Using the Box-Jenkins Method. J. Environ. Eng. Div. 1974, 100, 733–752. [CrossRef]
- 20. Ömer Faruk, D. A hybrid neural network and ARIMA model for water quality time series prediction. *Eng. Appl. Artif. Intell.* **2010**, 23, 586–594. [CrossRef]
- 21. Li, W.; Wu, H.; Zhu, N.; Jiang, Y.; Tan, J.; Guo, Y. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Inf. Process. Agric.* 2021, *8*, 185–193. [CrossRef]
- 22. Altunkaynak, A.; Özger, M.; Çakmakcı, M. Fuzzy logic modeling of the dissolved oxygen fluctuations in Golden Horn. *Ecol. Model.* **2005**, *189*, 436–446. [CrossRef]
- 23. Yin, L.; Fu, L.; Wu, H.; Xia, Q.; Jiang, Y.; Tan, J.; Guo, Y. Modeling dissolved oxygen in a crab pond. *Ecol. Model.* 2021, 440, 109385. [CrossRef]
- 24. Langendorf, R.E.; Lyubchich, V.; Testa, J.M.; Zhang, Q. Inferring Controls on Dissolved Oxygen Criterion Attainment in the Chesapeake Bay. ACS ES&T Water 2021, 1, 1665–1675. [CrossRef]
- 25. Gökçe, A. A mathematical study for chaotic dynamics of dissolved oxygen- phytoplankton interactions under environmental driving factors and time lag. *Chaos Solitons Fractals* **2021**, *151*, 111268. [CrossRef]
- 26. Ren, Q.; Wang, X.; Li, W.; Wei, Y.; An, D. Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network. *Aquac. Eng.* 2020, *90*, 102085. [CrossRef]
- Zhu, N.; Ji, X.; Tan, J.; Jiang, Y.; Guo, Y. Prediction of dissolved oxygen concentration in aquatic systems based on transfer learning. Comput. Electron. Agric. 2021, 180, 105888. [CrossRef]
- Wang, L.; Jiang, Y.; Qi, H. Marine Dissolved Oxygen Prediction with Tree Tuned Deep Neural Network. *IEEE Access* 2020, 8, 182431–182440. [CrossRef]
- Khan, U.T.; Valeo, C. Optimising Fuzzy Neural Network Architecture for Dissolved Oxygen Prediction and Risk Analysis. Water 2017, 9, 381. [CrossRef]

- Najah, A.; El-Shafie, A.; Karim, O.A.; El-Shafie, A.H. Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environ. Sci. Pollut. Res.* 2014, 21, 1658–1670. [CrossRef]
- Ji, X.; Shang, X.; Dahlgren, R.A.; Zhang, M. Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: A case study of Wen-Rui Tang River, China. *Environ. Sci. Pollut. Res.* 2017, 24, 16062–16076. [CrossRef]
- Liu, S.; Xu, L.; Li, D.; Li, Q.; Jiang, Y.; Tai, H.; Zeng, L. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. *Comput. Electron. Agric.* 2013, 95, 82–91. [CrossRef]
- Huan, J.; Li, M.; Xu, X.; Zhang, H.; Yang, B.; Jianming, J.; Shi, B. Multi-step prediction of dissolved oxygen in rivers based on random forest missing value imputation and attention mechanism coupled with recurrent neural network. *Water Supply* 2022, 22, 5480–5493. [CrossRef]
- 34. Huan, J.; Chen, B.; Xu, X.G.; Li, H.; Li, M.B.; Zhang, H. River Dissolved Oxygen Prediction Based on Random Forest and LSTM. *Appl. Eng. Agric.* 2021, *37*, 901–910. [CrossRef]
- Berkani, S.; Guermah, B.; Zakroum, M.; Ghogho, M. Spatio-Temporal Forecasting: A Survey of Data-Driven Models using Exogenous Data. *IEEE Access* 2023, 11, 75191–75214. [CrossRef]
- Zhang, J.; Xiao, F.; Li, A.; Ma, T.; Xu, K.; Zhang, H.; Yan, R.; Fang, X.; Li, Y.; Wang, D. Graph neural network-based spatio-temporal indoor environment prediction and optimal control for central air-conditioning systems. *Build. Environ.* 2023, 242, 110600. [CrossRef]
- Wu, Y.; Sun, L.; Sun, X.; Wang, B. A hybrid XGBoost-ISSA-LSTM model for accurate short-term and long-term dissolved oxygen prediction in ponds. *Environ. Sci. Pollut. Res.* 2022, 29, 18142–18159. [CrossRef]
- Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* 2020, 34, 415–433. [CrossRef]
- 39. Yang, Y.; Xiong, Q.; Wu, C.; Zou, Q.; Yu, Y.; Yi, H.; Gao, M. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environ. Sci. Pollut. Res.* 2021, *28*, 55129–55139. [CrossRef]
- Haq, K.P.R.A.; Harigovindan, V.P. Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models. *IEEE Access* 2022, 10, 60078–60098. [CrossRef]
- Jiang, F.; Tao, W.; Liu, S.; Ren, J.; Guo, X.; Zhao, D. An End-to-End Compression Framework Based on Convolutional Neural Networks. *IEEE Trans. Circuits Syst. Video Technol.* 2018, 28, 3007–3018. [CrossRef]
- Shea, A.O.; Lightbody, G.; Boylan, G.; Temko, A. Neonatal seizure detection using convolutional neural networks. In Proceedings of the 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), Tokyo, Japan, 25–28 September 2017; pp. 1–6.
- Manickathan, L.; Mucignat, C.; Lunati, I. Kinematic training of convolutional neural networks for particle image velocimetry. *Meas. Sci. Technol.* 2022, 33, 124006. [CrossRef]
- 44. Fanta, H.; Shao, Z.; Ma, L. SiTGRU: Single-Tunnelled Gated Recurrent Unit for Abnormality Detection. *Inf. Sci.* 2020, 524, 15–32. [CrossRef]
- 45. Kumari, P.; Toshniwal, D. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Appl. Energy* **2021**, *295*, 117061. [CrossRef]
- Yibei, L.; Peishun, L.; Xuefang, W.; Xueqing, Z.; Zifei, Q. A study on water quality prediction by a hybrid dual channel CNN-LSTM model with attention mechanism. In Proceedings of the International Conference on Smart Transportation and City Engineering, Chongqing, China, 6–8 August 2021; p. 1205035.
- 47. Lu, Y.; Wang, W.; Hu, X.; Xu, P.; Zhou, S.; Cai, M. Vehicle Trajectory Prediction in Connected Environments via Heterogeneous Context-Aware Graph Convolutional Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *24*, 8452–8464. [CrossRef]
- Buyukdemircioglu, M.; Kocaman, S. Reconstruction and Efficient Visualization of Heterogeneous 3D City Models. *Remote Sens.* 2020, 12, 2128. [CrossRef]
- Kokkinos, K.; Karayannis, V.; Nathanail, E.; Moustakas, K. A comparative analysis of Statistical and Computational Intelligence methodologies for the prediction of traffic-induced fine particulate matter and NO<sub>2</sub>. J. Clean. Prod. 2021, 328, 129500. [CrossRef]
- Rixen, T.; Cowie, G.; Gaye, B.; Goes, J.; do Rosário Gomes, H.; Hood, R.R.; Lachkar, Z.; Schmidt, H.; Segschneider, J.; Singh, A. Reviews and syntheses: Present, past, and future of the oxygen minimum zone in the northern Indian Ocean. *Biogeosciences* 2020, 17, 6051–6080. [CrossRef]
- Pandey, K.; Patel, S. Deep Learning with Convolutional Neural Networks: From Theory to Practice. In Proceedings of the 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 11–13 April 2023; pp. 1217–1224.
- 52. Ayesha Jasmin, S.; Ramesh, P.; Tanveer, M. An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques. *Expert Syst. Appl.* **2022**, *199*, 117160. [CrossRef]
- Ghosh, A.; Jana, N.D. Artificial Bee Colony Optimization based Optimal Convolutional Neural Network Architecture Design. In Proceedings of the 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 24–26 November 2022; pp. 1–7.
- 54. Deng, Y. Recommender Systems Based on Graph Embedding Techniques: A Review. IEEE Access 2022, 10, 51587–51633. [CrossRef]
- 55. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. arXiv 2017, arXiv:1706.03762.
- 56. Ferrari, I.; Frisoni, G.; Italiani, P.; Moro, G.; Sartori, C. Comprehensive Analysis of Knowledge Graph Embedding Techniques Benchmarked on Link Prediction. *Electronics* **2022**, *11*, 3866. [CrossRef]

- 57. Xu, M. Understanding Graph Embedding Methods and Their Applications. SIAM Rev. 2021, 63, 825–853. [CrossRef]
- Liu, Y.; Zeng, K.; Wang, H.; Song, X.; Zhou, B. Content Matters: A GNN-Based Model Combined with Text Semantics for Social Network Cascade Prediction. In Proceedings of the Advances in Knowledge Discovery and Data Mining, Cham, Switzerland; 2021; pp. 728–740.
- 59. Réau, M.; Renaud, N.; Xue, L.C.; Bonvin, A.M.J.J. DeepRank-GNN: A graph neural network framework to learn patterns in protein–protein interfaces. *Bioinformatics* **2023**, *39*, btac759. [CrossRef]
- Rocco, J.D.; Sipio, C.D.; Ruscio, D.D.; Nguyen, P.T. A GNN-based Recommender System to Assist the Specification of Metamodels and Models. In Proceedings of the 2021 ACM/IEEE 24th International Conference on Model Driven Engineering Languages and Systems (MODELS), Fukuoka, Japan, 10–15 October 2021; pp. 70–81.
- 61. Kalyan, K.S.; Rajasekharan, A.; Sangeetha, S. AMMU: A survey of transformer-based biomedical pretrained language models. *J. Biomed. Inform.* **2022**, *126*, 103982. [CrossRef]
- Jin, Y.; Hou, L.; Chen, Y. A Time Series Transformer based method for the rotating machinery fault diagnosis. *Neurocomputing* 2022, 494, 379–395. [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.