Article

New Graph-Based and Transformers Deep Learning Models for River Dissolved Oxygen Forecasting

Paulo Alexandre Costa Rocha [*] , Victor Oliveira Santos , Jesse Van Griensven Thé , Bahram Gharabaghi [*]

Article

New Graph-Based and Transformers Deep Learning Models for River Dissolved Oxygen Forecasting

Paulo Alexandre Costa Rocha [1,2,*], Victor Oliveira Santos [1], Jesse Van Griensven Thé [1,3] and Bahram Gharabaghi [1]

[1] School of Engineering, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada; volive04@uoguelph.ca (V.O.S.)

[2] Mechanical Engineering Department, Technology Center, Federal University of Ceará, Fortaleza 60020-181, CE, Brazil

[3] Lakes Environmental Research Inc., 170 Columbia St W, Waterloo, ON N2L 3L3, Canada; jesse.the@weblakes.com (J.V.G.T.)

* Correspondence: Correspondence: pcostaro@uoguelph.ca (P.A.C.R.); bgharaba@uoguelph.ca (B.G.)

Abstract: An important indicator of human-related pollution in watersheds is dissolved oxygen (DO). The DO is highly dependent on both space and time characteristics of the watershed and is directly linked to eutrophication, which impairs the development of both the aquatic fauna and flora, also negatively impacting the water quality. Aspiring to reach a more accurate and precise forecasting approach to predict levels of DO, the present work proposes new graph-based and transformer-based deep learning models. The models were trained and validated for the Credit River Watershed, and the results were compared with both benchmarking and literature-found approaches. The proposed Graph Neural Network Sample and Aggregate (GNN-SAGE) model was the best-performing approach, reaching coefficient of determination (R^2) and Root Mean Squared Error (RMSE) values of 97% and 0.34 ppm, respectively. The findings from the Shapley additive explanations (SHAP) indicated that the GNN-SAGE benefited from spatiotemporal information from the surrounding stations, improving the model's results, and that temperature is a major input attribute for determining future DO levels. The results established that the proposed GNN-SAGE model stands as a state-of-the-art solution for DO forecasting, with potential for real-time water quality applications.

Keywords: pollution; dissolved oxygen; Credit River; machine learning; graph neural networks; SHAP analysis

1. Introduction

In the current Anthropocene Era [1], the crescent human occupation in regions close to rivers and lakes has negatively affected the watershed aquatic ecosystems, and the water quality [2,3]. Residues coming from locations within the watershed are major sources of pollutants, which may reach the waterbody through discharged runoff, thus disrupting the aquatic biota's physical, chemical, and biological conditions [4,5]. Monitoring the river water quality is important to keep track of the river's condition, helping develop management strategies aimed at the river's preservation and conservation [6]. One major indicator of the river's quality is the Dissolved Oxygen (DO) concentration. The DO is highly variable in both time and space, under the constant influence of atmospheric pressure, water temperature, turbidity, and many more [7,8]. Disturbs on DO concentrations are related to an increase in eutrophication, which reduces the total DO in the watershed, impairing aquatic life and human health, leading to economic losses [9–12]. Given the major role of DO in maintaining the health of aquatic ecosystems, it is crucial to forecast and monitor its level with high accuracy. This will help mitigate the impacts of anthropogenic changes on the aquatic environment and will also facilitate the development of decision-making strategies by the stakeholders, preventing further degradation of the environment.

To model and assess DO concentration in a watersheds, different paths were proposed. In recent years, thanks to the progress in hardware development [13,14], data-driven approaches like Machine Learning (ML) have been drawing attention from researchers due to their improved computational performance and generalization capabilities when compared to numerical models, also being able to identify complex non-linear associations in the dataset without prior coding [15,16]. When used in time-series forecasting applications, the ML paradigm reached excellent results in different knowledge fields [17–20]. Extensive literature can be found for hydrological studies using ML tools. In [21], the authors verified different ML models for DO forecasting for the year 2003 on the Danube River, in Hungary. Their proposed non-linear approaches proved to achieve better results when compared to other linear models, reaching the best Root Mean Squared Error (RMSE) of 1.36 mg/L when the proposed general regression neural networks were implemented. Decision tree-based models were evaluated in work [22]. The authors merged the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) with tree-structured models to forecast several water quality variables, including DO concentrations. For short-term predictions on the Tualatin River, USA, their results showcased that the best strategy to determine future DO levels was achieved by combining CEEMDAN and random forests (RF), reached the oprtimal RMSE value of 0.10 mg/L. Hybrid models based on wavelet transform (WT) and ML structures were studied in [23]. There, the authors assessed different forecasting approaches to determine future DO concentrations in Dongjiang River, in China, for 1 to 5 days of leading time. Among the evaluated structures, the best performing one for 1 day ahead forecasting was the combination of WT and Artificial Neural Networks (ANN), which achieved RMSE and coefficient of determination (R^2) of 0.499 mg/L and 96.1%, while for the 5 days ahead scenario the merged model of WT and multiple linear regression (MLR) reached the best outcomes, with 1.113 mg/L for RMSE and 78.8% for R^2.

Deep learning (DL) models are an advanced approach to determining future DO concentrations. Dissolved Oxygen estimation is a spatiotemporal problem where recurrent neural networks and their variations are popular among researchers in this field. In [24], authors used remote sensing data to estimate future DO concentrations using prior DO information in the Rawal watershed, in Pakistan. The best result for future DO estimation was returned by the Bi-directional Long Short-Term Memory (Bi-LSTM) paradigm, with an RMSE equal to 0.199 mg/L. Hybrid configuration merging the LSTM and the Convolutional Neural Network (CNN) paradigm was evaluated in [25], to forecast short-term DO in Small Prespa Lake, Greece. Compared to the standalone models and conventional ML paradigms such as Decision Trees (DT) and Support Vector Machine (SVM), the hybrid configuration surpassed all of them, reaching an RMSE of 0.518 mg/L. Other application using CEEMDAN together with DL models was proposed in work [26]. In this paper, the authors compared several DL models to real-time forecasting of DO concentrations in Xin'anjiang River, China. They compared the standalone DL appraoches CNN, LSTM, the hybrid CNN-LSTM, and combinations of these models with the CEEMDAN pre-processing. The authors discovered that the CEEMDAN-CNN-LSTM hybrid configuration was able to provide the best outcomes for different multi-step ahead forecasting for DO, with the best RMSE value of 0.26 mg/L for 4-h ahead forecasting. In recent years, graph neural networks (GNN) models have been implemented to time-series forecasting within the DL field, reaching cutting-edge results [27–29]. However, little is known about implementing such a paradigm in the water quality research area, with few works addressing this subject [30–33].

As the present bibliographic review elucidates, ML applications in hydrological and environmental situations constitute a vast field of study. Aiming to deepen the understanding of ML, specifically addressing the graph-based approaches, the present work aims to contribute to filling the knowledge gap regarding these models when applied to DO forecasting. To this end, historical data comprising the years 2016 to 2020 for Credit River, Ontario, were collected and implemented to the Deep Neural Network Transformer (DNN-Transformer) and the Graph Neural Network Sample and Aggregate (GNN-SAGE) to determine future levels of DO on the river, a highly non-conservative substance. The present study aims to contribute to the field by:

1. Developing a cutting-edge model to predict DO concentration with elevated precision and accuracy.

2. Verifying the temporal effect over the estimated DO concentration, a highly non-conservative substance, by implementing different time lags for different predictive models, namely XGBoost, DNN-Transformer, and the proposed GNN-SAGE.

3. Conducting a Shapley additive explanations (SHAP) analysis to assess the significance of different input variables, allowing meaningful inputs over the models' forecasting and its functioning.

4. Facilitating the development of a real-time water quality tracking system for urban rivers, aiding the elaboration of risk management strategies and environmental policies.

This work is divided as follows: in Section 2, the case study of the Credit River is presented. The methodology used for each implemented model and the SHAP analysis is presented in Section 3. Section 4 holds the results for the forecasting models, which are then discussed in Section 5. Section 6 closes our work with a conclusion.

2. Case Study

The Credit River is a major river in Southern Ontario, Canada, part of the Toronto Greater Area [34]. Starting at Orangeville, the Credit River stretches for 90 km, before finally reaching Lake Ontario [35,36]. The river has a mainly rural configuration in its upper and middle areas, having significant forest coverage [37]. As it approaches Lake Ontario, the urbanization of the river's area increases [6]. Urban development in the river area, however, endangers aquatic life and the people who rely on the river, increasing the risk of floods [35,38–40]. Figure 1 shows the location of the studied site in Ontario province. The red point in the map represents the station where the dissolved oxygen is predicted, namely "Credit River @ MGCC". The green points are the used neighbouring monitoring stations, which provide spatiotemporal information to the forecasting model.

This study developed the forecasting models using historical data from 2016 to 2020. This data included weather attributes and physical-chemical information collected from monitoring stations located throughout the river's watershed, as shown in Figure 1. The dataset, originally with a temporal resolution of 15 minutes, was resampled to 1-hour intervals. This adjustment involved calculating the average value for each variable, except for precipitation, which was summed. These processed data were then used as input for the models under evaluation. The data correlogram is presented in Figure 2, where the target attribute is the dissolved oxygen (first row and first column).
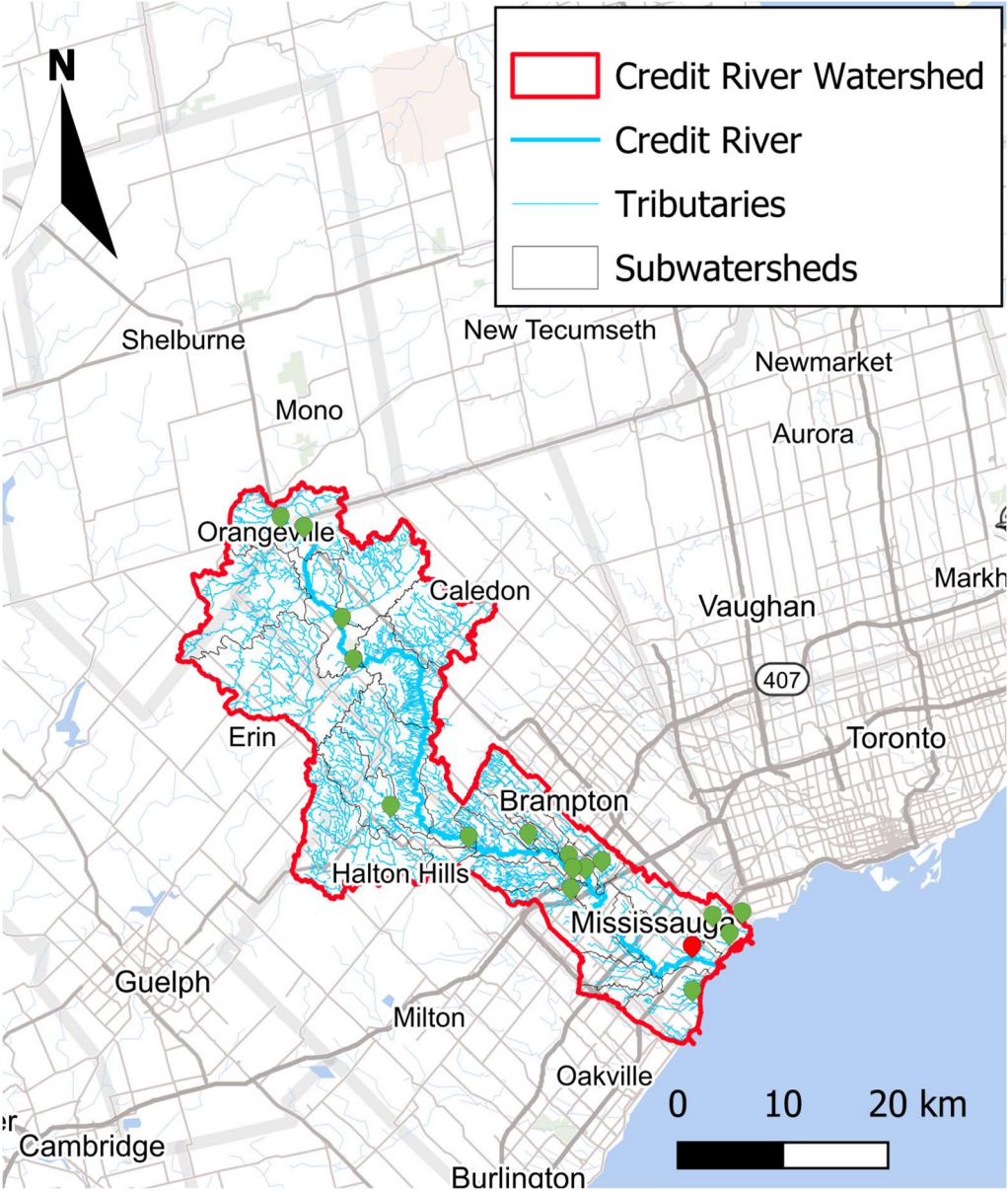
Figure 1. Location of the Credit River in Ontario Province. The points showcase the position of the measurement stations along the river's course.
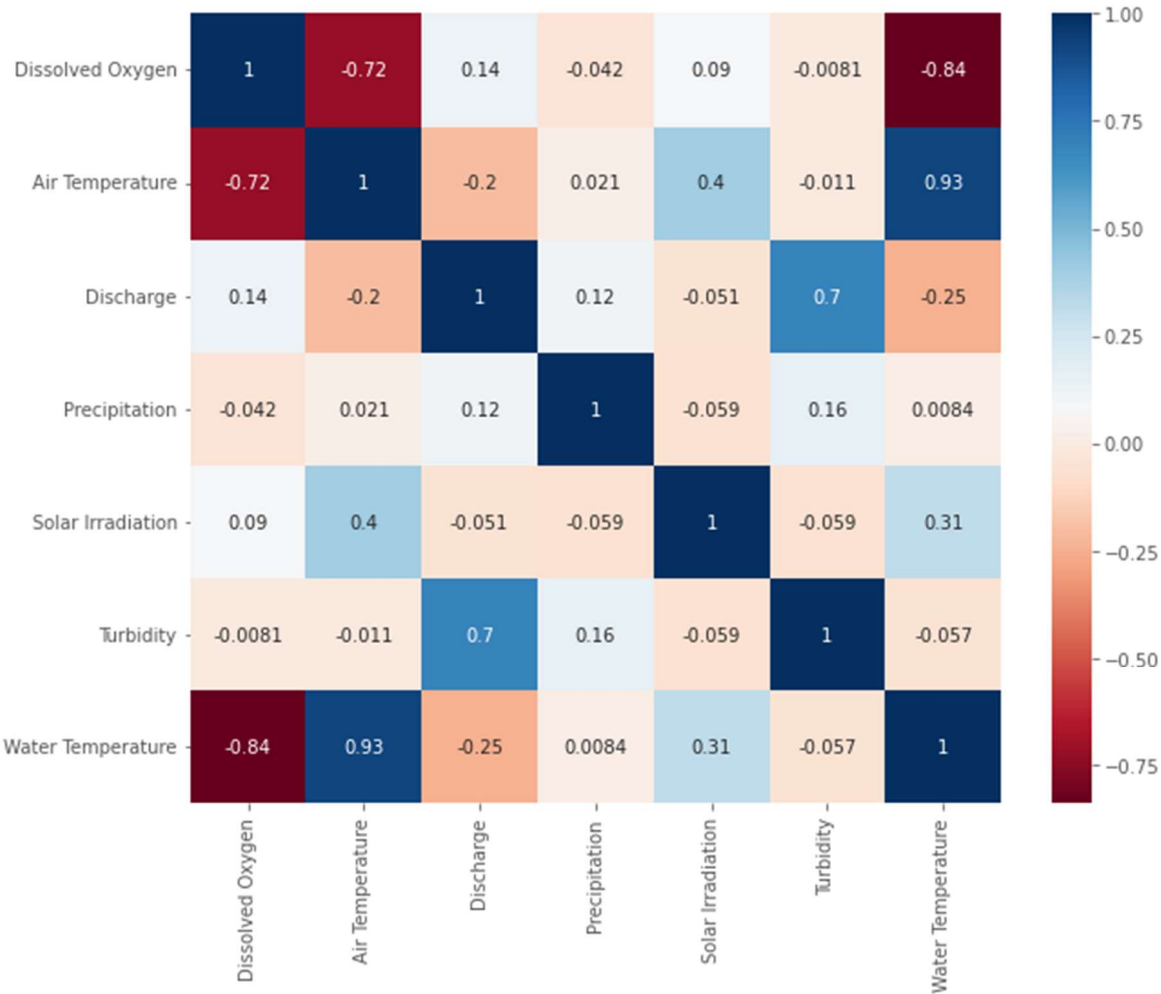
Figure 2. Correlation matrix for the Credit River variables used for dissolved oxygen concentration forecasting.

In Figure 2, the correlation matrix is depicted, where blue colors indicate positive correlation and red color shows negative correlation. The used input data presents a strong negative correlation between water temperature and air temperature, an expected behaviour for DO since higher temperatures impair gas solubility in water [41–43]. A weak positive correlation exists between the discharge input, by disturbing the water body, and solar irradiation, by influencing the aquatic photosynthesis, both impacting the DO levels in the river. The remaining precipitation and turbidity showed a very small correlation to the dissolved oxygen attribute, indicating that they may not affect the DO levels in the water body. Also, the negative value of these variables indicates that they may reduce the photosynthetic process by the aquatic flora [44,45].

It is important to note that the occurrence of high correlation values in different input attributes does not necessarily positively impact the forecasting result. This may introduce high variance into the predictive model due to collinearity, sometimes damaging its performance [46,47]. However, incorporating input variables with low correlation values into the forecasting model may be advantageous. These inputs may contain valuable spatiotemporal information that can help reveal the connection between the input and output parameters, ultimately improving the model's predictions [48].

3. Methodology

3.1. Benchmarking Models

The achievements of the assessed models were evaluated against the results of two benchmarking models, namely the persistence and eXtreme Gradient Boosting (XGBoost). The persistence model is a simple benchmarking widely used to evaluate any forecasting model. It simply states that the current value of the target variable is the same as the previously measured one. This approach has good performance on forecasting values in short time windows, but as the forecasting horizon increases, it can no longer capture the complex non-linear behaviour underneath the dataset nor external factors influencing future values, quickly deteriorating its results [49,50].

XGBoost is a tree-based ML model, improving the random forests approach. Based on the bagging sampling, the XGBoost algorithm generates smaller tree models. It later combines them into a unique bigger one, reducing the final variance and, therefore the risk of overfitting [51–53]. The XGBoost model performs remarkably well in handling missing values in the dataset, and by being scalable to different applications [54], it is successfully applied in various research fields [55–57].

3.2. DNN-Transformer Model

The evaluated DNN-Transformer model utilizes a deep learning technique that incorporates a self-attention mechanism. This enables the model to concentrate on important features of the input information, improving its decision-making capabilities [58]. On its original configuration [59], the transformer architecture determines how the data attends to each one of its elements by implementing positional encoding together with multi-head attention, later being processed by a feed-forward structure that will output the processed data by the encoder. The generated information from the encoder is later used in the decoder structure, where a multi-head structurea multi-head structure will again process it will again process it before being fed into another feed-forward internal model, which will output the transformer result.

For the present work, the transformer architecture was adapted so it can be applied to a time-series forecasting model. The used DNN-Transformer architecture is depicted in Figure 3.
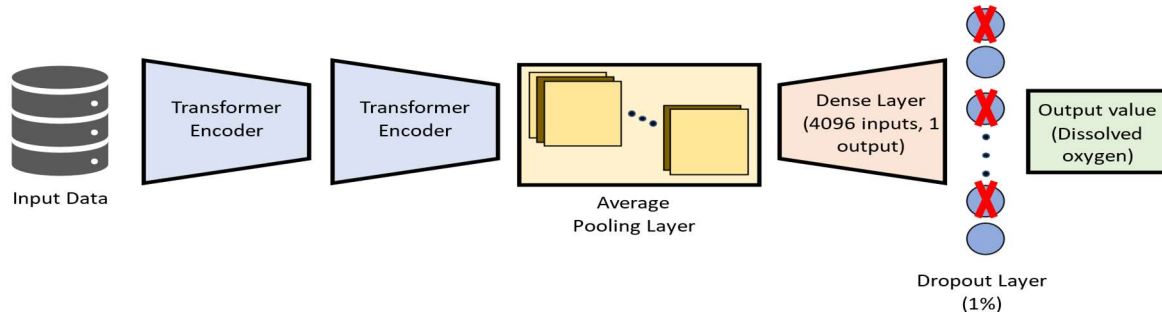

Figure 3. Applied DNN-Transformer model structure.

In the used architecture, the input data passes through two transformer structures. Then, its output is processed by an average pooling layer, which connects to a dense layer. The dropout layer [60] is then used to avoid overfitting by relying on an ensemble of different models after the training phase, thus reducing the model's variance [52,61]. Finally, the forecasted value for dissolved oxygen is output by the DNN-Transformer model.

3.3. GNN-SAGE Model

The Graph Neural Network Sample and Aggregate (GNN-SAGE) is a cutting-edge model which applies both graph theory and deep learning techniques. Unlike traditional ML models, the graph neural networks approach considers graph-structured data naturally, retaining relevant spatiotemporal information ruling the behavior of input variables and target variables, boosting time-series forecasting models [62,63]. For the SAGE paradigm, the model can generalize unknown data by sampling a constant set of nodes, later aggregating them using an aggregator function [64–66], achieving excellence in extracting spatiotemporal information underneath the data [28,29]. The structure used in the present work is depicted in Figure 4.
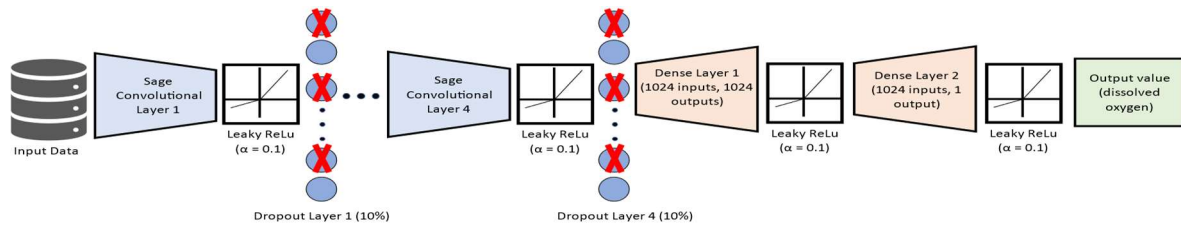
Figure 4. Applied GNN-SAGE model structure.

Figure 4 feeds the input spatiotemporal data into the GNN-SAGE model. First, it passes through a succession of graph convolutional layers, followed by the Leaky ReLU function [67,68] and a subsequent dropout layer. After four convolution processes, the data is fed to a dense layer containing 1024 inputs and 1024 outputs, followed again by the Leaky ReLu activation function and a final dense layer containing 1024 inputs and 1 output. Finally, future dissolved oxygen values are then generated by the GNN-SAGE.

One key advantage of graph-based architectures, like GNN-SAGE, over traditional DL approaches is their natural ability to process multi-spatiotemporal data from surrounding monitoring stations. This enables them to effectively uncover the inherent connection among input and target variables when implemented in predictive applications [48,62,63]. The capacity of GNN-SAGE to extract spatiotemporal information from data is relevant to this study, given that DO levels depend on the dataset's temporal and spatial features.

3.3. SHAP Analysis

Despite the excellent results achieved by ML and DL models in recent studies, these models still lack interpretability, posing as a major hindrance to a full understanding and interpretation of how they work [69–71]. Aiming to solve this problem, the Shapley Additive Explanations (SHAP) statistic was developed to be a valuable tool for better understanding the black box behaviour of ML applications. This game theory-based technique provides a deeper understanding of ML models by examining the relationship among input attributes and the resulting outcome. It does so by combining additive feature attribution in terms of each input variable's significance, correlation, and impact on the final prediction, resulting in a better understanding of how complex models work [72].

4. Results

4.1. Evaluation of different time lags over the model's performance

The performances of the DNN-Transformer and GNN-SAGE approaches were compared to the benchmarking ones for different number of time lags. In the present work, the time lag consists of a subseries of the original time series dataset, representing past information in hours, that is fed to the model. That is, being x_i the attribute value taken in the i-th hour, its time lag is defined as a subgroup of the attribute x, where $\{x_{i-1}, x_{i-2}, …, x_{i-n}\}$ contains up to n time lags. The results are presented in Figure 5.
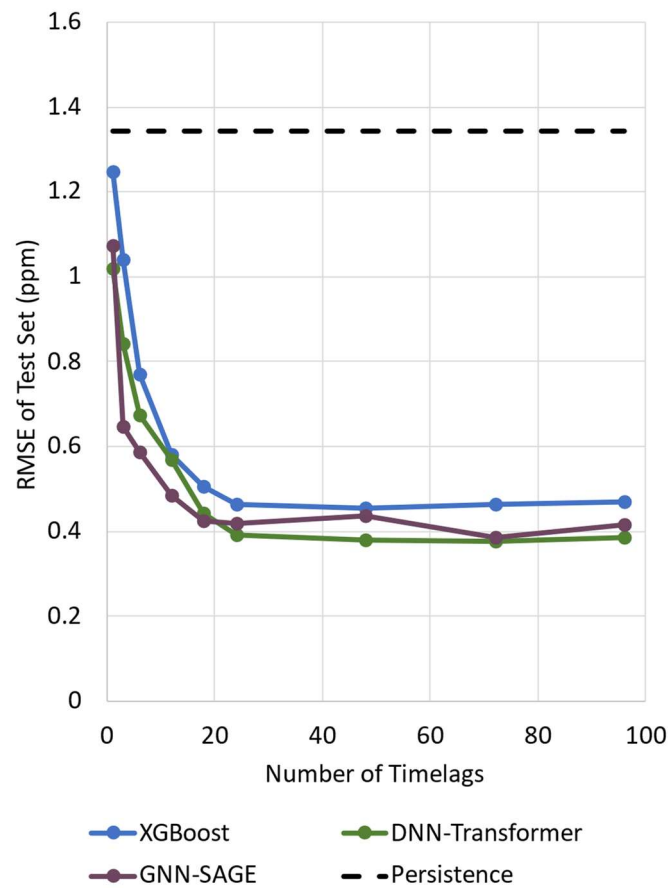
Figure 5. Impact of different number of time lags for dissolved oxygen forecasting.

Figure 5 illustrates the evaluated models' performance using different time lags by using only dissolved oxygen as the input parameter. For the assessed time lag values, all the models could surpass the persistence and XGBoost benchmarking, presenting similar behaviour: a steep decrease in the RMSE value, followed by convergence after 20 time lags. The deep learning paradigms, DNN-Transformers, and GNN-SAGE, benefited more as time lags increased. Both XGBoost and DNN-Transformers presented more stable outcomes when compared with GNN-SAGE, which demonstrated a small variation in its results.

For the evaluated models, the best results were achieved by the DNN-Transformers for 48 time lags (RMSE = 0.38 ppm). The proposed GNN-SAGE approach was the second best-performing model, reaching the optimal RMSE value of 0.39 for 72-time lags, a comparable value for the DNN-Transformer. The DNN-Transformer model improved dissolved oxygen forecasting by 72% compared to the persistence model, and by 17% in comparison with XGBoost. For the GNN-SAGE, the improvements were 71% and 15%, when compared against the persistence and XGBoost models, respectively. Finally, using the results depicted in Figure 5, we implemented 72 time lags for the input attributes, i.e., data from up to the previous 72 hours for the variables displayed in Figure 2, including dissolved oxygen concentrations, to determine future DO levels using GNN-SAGE, while DNN-Transformer was implemented using 48 time lags.

4.2. Results of dissolved oxygen for 6 h ahead

To determine the optimal configuration for forecasting dissolved oxygen levels, various input variables were tested. Using a step-by-step approach [48] for the considered parameters, the models' results were initially evaluated using just past dissolved oxygen information. Additional input parameters were then added to the models, and the resulting forecasts were analyzed. If the addition

of an attribute did not improve the model's forecasting performance, leading to a greater error than previous configurations, that attribute was removed. This process was repeated for all models until all attributes depicted in Figure 2 were assessed, considering a 6 h forecasting horizon. The Figure 6 depicts the results for this approach, being the x-axis shows the used input variables combination and, in the y-axis, their respective RMSE values in ppm. In Figure 6, the lighter the column colour, the lower the RMSE value.
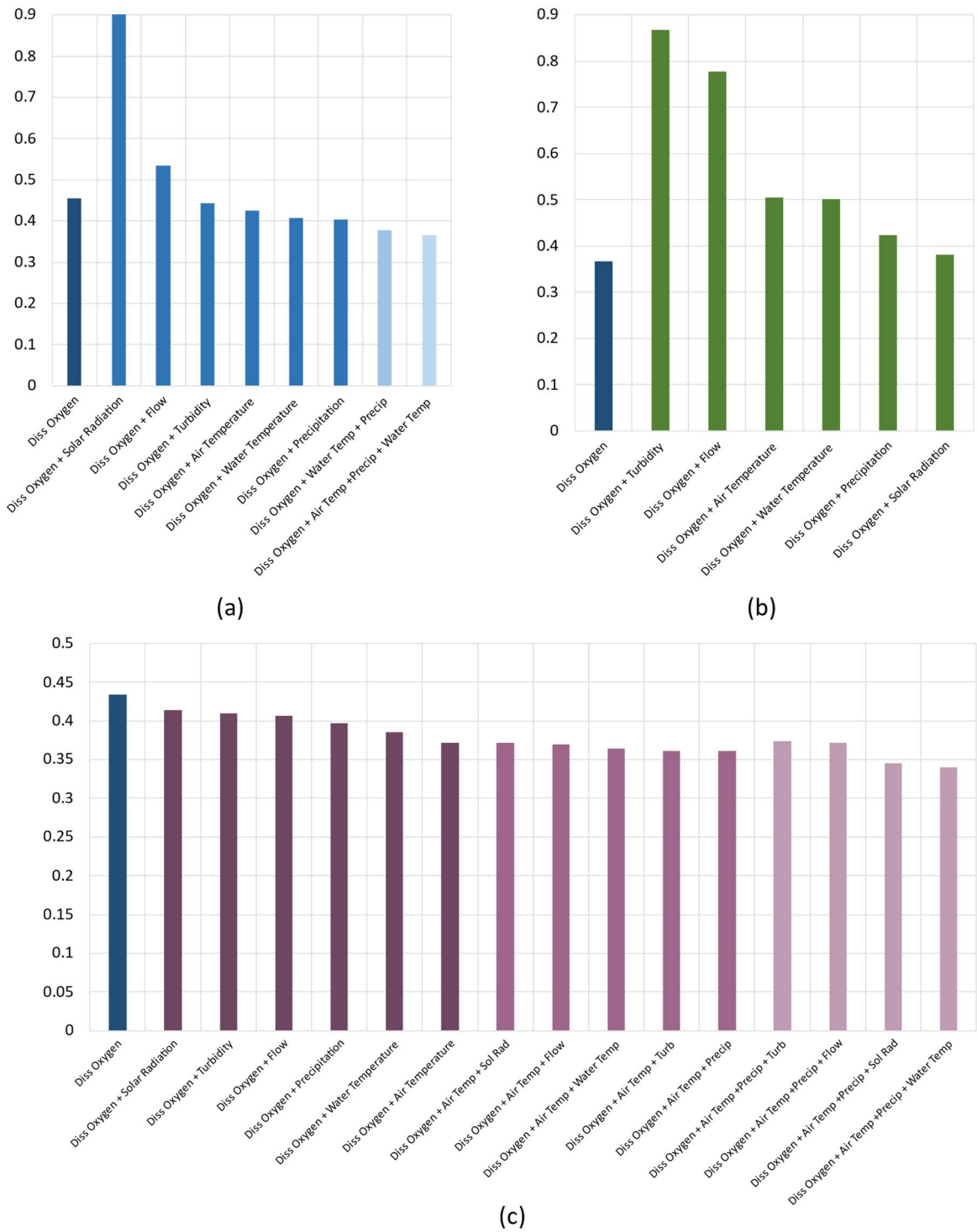

Figure 6. The effect of different input parameters for dissolved oxygen 6 h ahead estimation using (a) XGBoost, (b) DNN-Transformer, and (c) GNN-SAGE models.

Figure 6 presents the different forecasting performances for each one of the evaluated models. In panel Figure 6 (a), it is possible to observe that the XGBoost approach reached improved predictive results as more input variables were added. Its best configuration was set as using Dissolved Oxygen, Air Temperature, Precipitation, and Water Temperature, achieving an RMSE value of 0.37 ppm. Interestingly, the DL transformer-based model's best configuration reached the same error value as the XGBoost. Analyzing Figure 6 (b), using only dissolved oxygen as an input attribute, the DNN-Transformer achieved an RMSE outcome of 0.37 ppm. The model did not need to use any exogenous variables to achieve its best result. Contrarywise, the last panel indicates that the GNN-SAGE also benefited from increased input parameters. In fact, it was the best-performing model, reaching an RMSE of 0.34 ppm when fed with Dissolved Oxygen, Air Temperature, Precipitation, and Water Temperature. The GNN-SAGE model enhanced the DO forecasting performance by 8% when compared to both XGBoost and DNN-Transformer.

Further investigation on the forecasting capacity of XGBoost, DNN-Transformer, and GNN-SAGE can be evaluated by analyzing Figures 7 to 9, which contain a scatter plot and marginal distributions for the real measured dissolved oxygen and the predicted values by the models, in ppm.
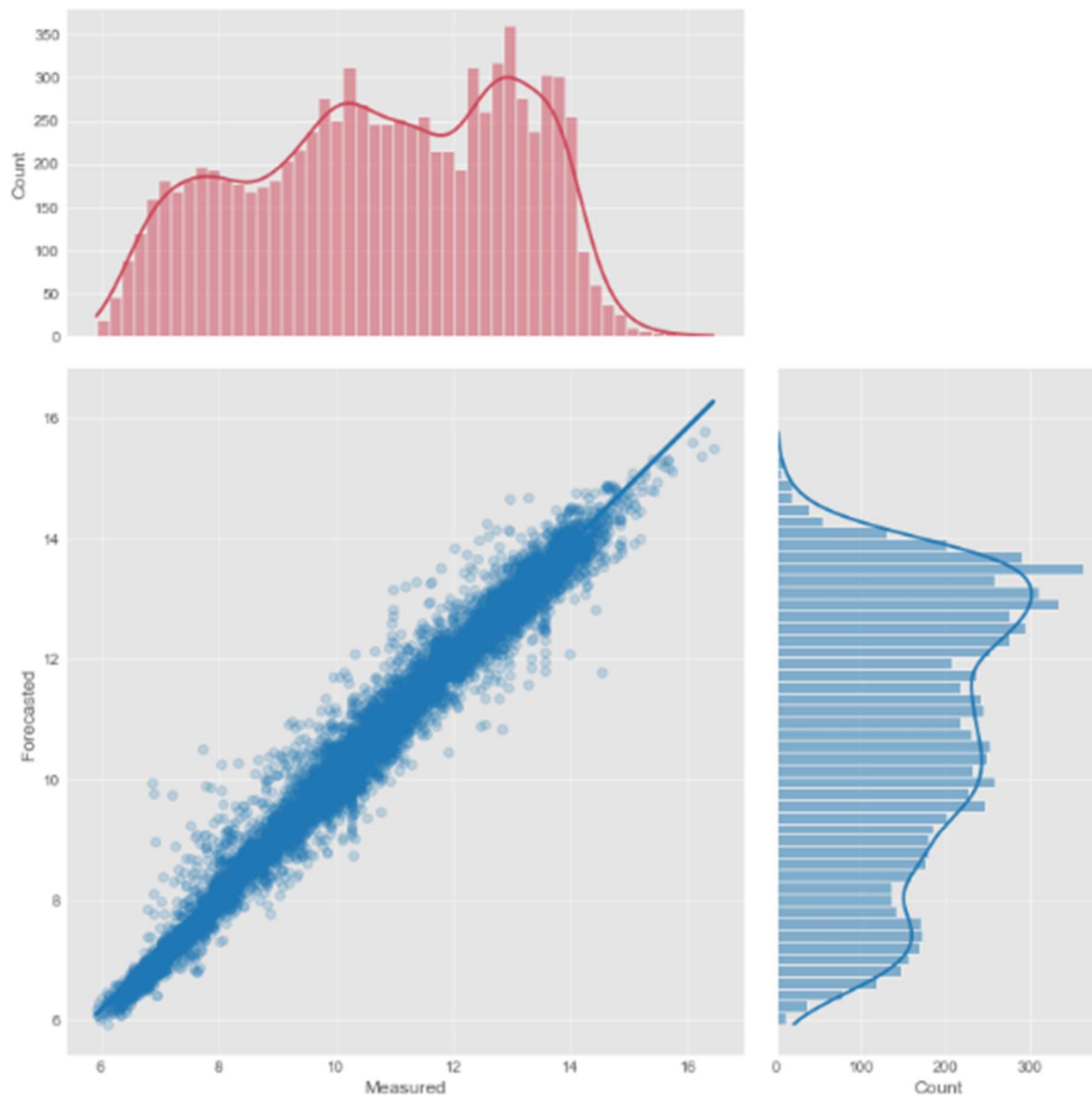

Figure 7. Scatter plot illustrating the estimated and observed dissolved oxygen levels for a forecast window of 6 h using XGBoost.
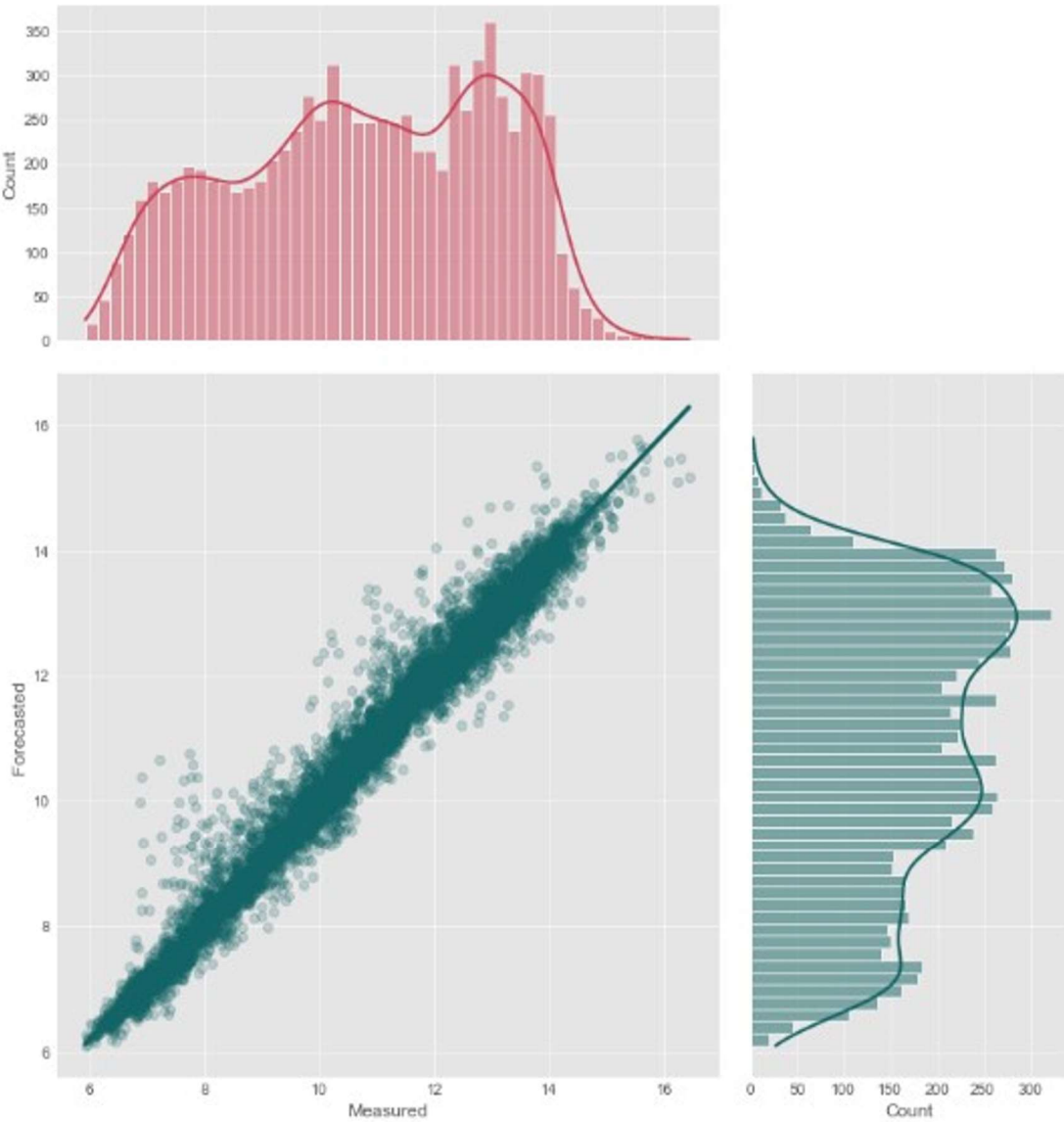
Figure 8. Scatter plot illustrating the estimated and observed dissolved oxygen levels for a forecast window of 6 h using DNN-Transformer.
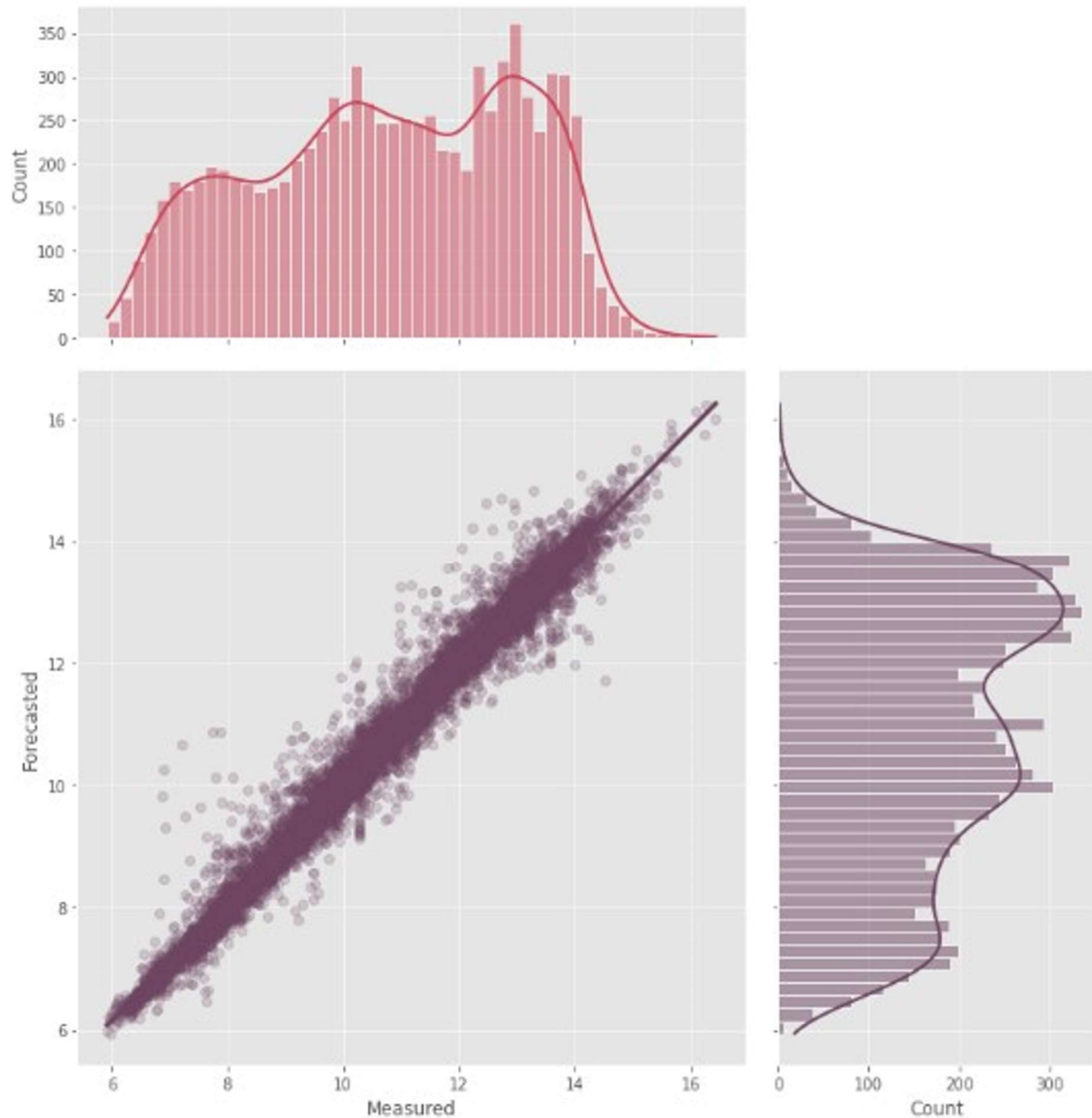
Figure 9. Scatter plot illustrating the estimated and observed dissolved oxygen levels for a forecast window of 6 h using GNN-SAGE.

Figures 7 to 9 present very similar results for the three evaluated models. They have well-aligned scatter points with the regression line, and very similar data distribution, as indicated by the histograms. However, the GNN-SAGE model has superior performance over the tree-based and transformer-based approaches, reaching an excellent coefficient of determination of 97.6%. Figures 10 to 12 show the models' results for 30 days, comparing them with real measured data for dissolved oxygen.
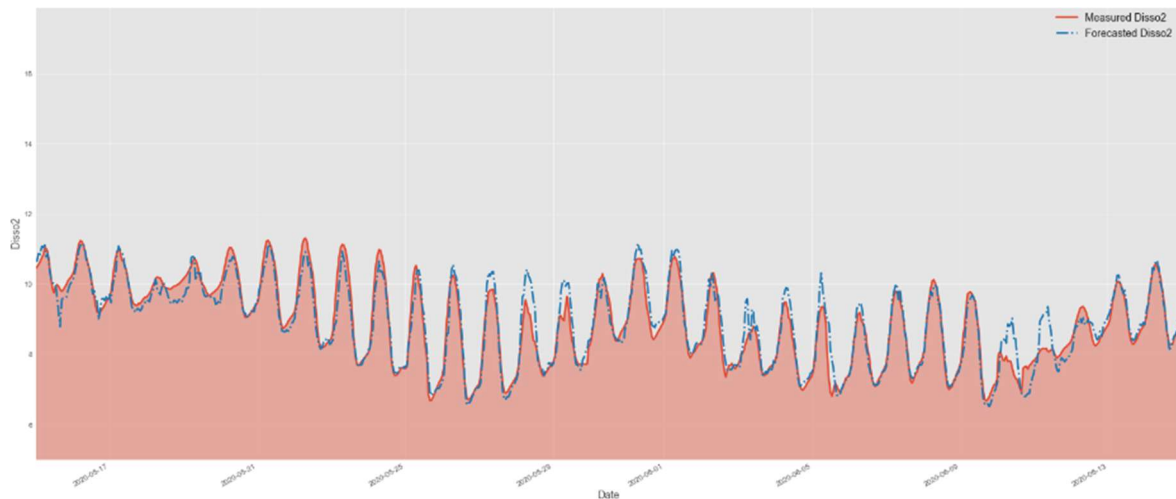
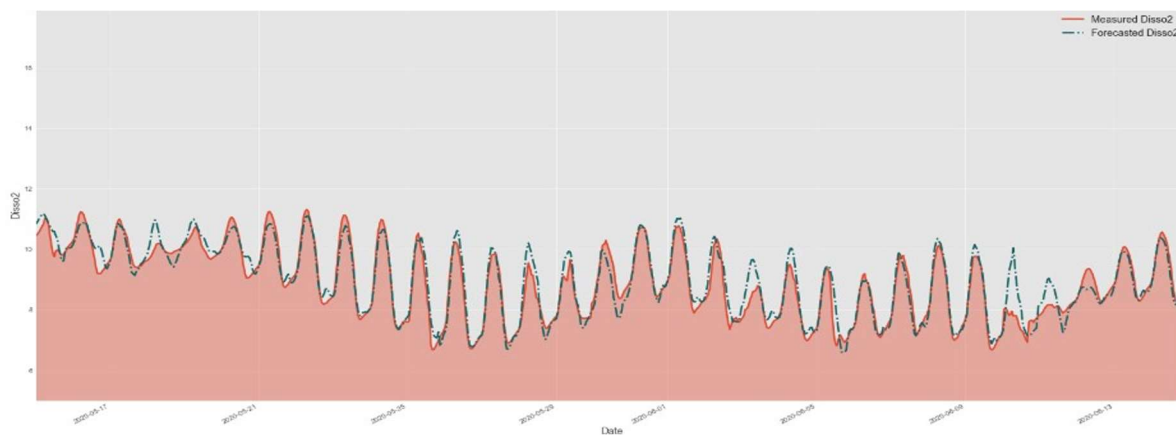**Figure 10.** Comparison of test data and forecasted data by XGBoost model.

Figure 11. Comparison of test data and forecasted data by DNN-Transformer model.
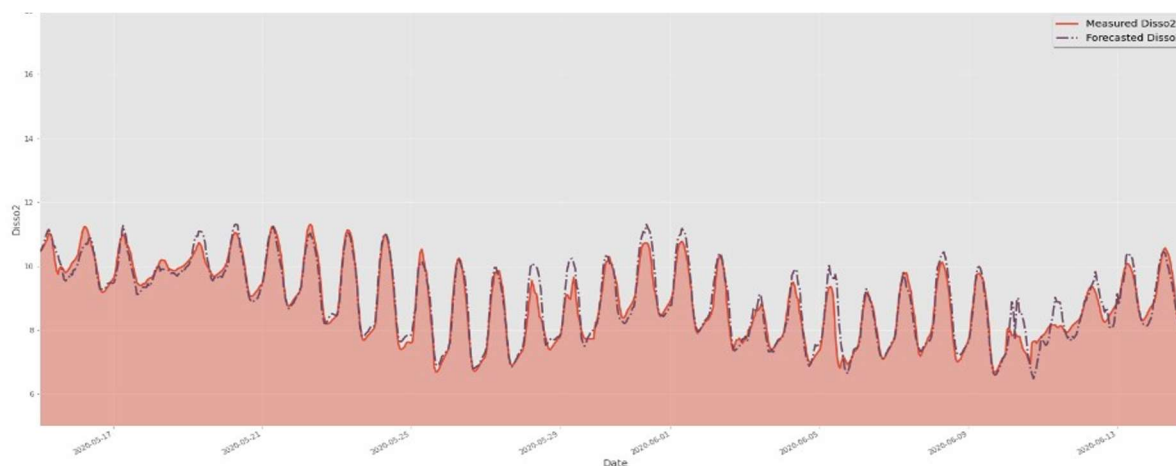
Figure 12. Comparison of test data and forecasted data by GNN-SAGE model.

In Figures 10 to 12, comparing the evaluated models' results for the same assessed period, we can observe that the graph-based structure has a superior performance in identifying the peak concentrations of dissolved oxygen, without any lagging in its results. Comparing the XGBoost to the GNN-SAGE, we notice that the graph-based approach has superior performance in identifying DO peaks, closely following their shape with a marginal overestimation. Comparing the transformer-based and the GNN-SAGE models, we can notice that the DNN-Transformer model tends to maintain

the historical behaviour of the DO, as seen in the fifth peak from right to left in Figure 11. This is an expected behaviour from the transformer approach since it determines the most likely outcome based on the previous sequential data [59]. Contrariwise, the GNN-SAGE can aggregate the influence of the spatiotemporal relationships of the input parameters, deeming it sensitive to changes in the DO concentration and thus identifying the peaks within the assessed period, as seen in the same fifth peak from right to left in Figure 12.

5. Analysis of Results

5.1. Analysis of Results of Dissolved Oxygen for 6 h ahead

When applied to the proposed scenario, the graph-based model GNN-SAGE presented the best performance, providing cutting-edge results for DO forecasting, a highly non-conservative substance, for 6-h ahead forecasting. By using Dissolved Oxygen, Air Temperature, Precipitation, and Water Temperature as input attributes, the proposed GNN-SAGE model showed that, when compared to XGBoost and DNN-Transformer, the graph structure could better model and analyze DO levels. The GNN-SAGE could satisfactorily extract spatiotemporal data from the study station and its surroundings, substantially improving the DO forecasting results, reaching R^2 of 97.6% and RMSE of 0.34 ppm, representing improved DO forecasting by 8% compared to both XGBoost and DNN-Transformer RMSE. The superior performance of the GNN-SAGE for hydrological applications has been attested in previous studies [28,48].

The proposed model also showed better tracking of DO peaks during 4 weeks, as presented in Figures 7 to 9. In contrast to the transformer-based approach, the GNN-SAGE model proved to be more sensitive to the spatiotemporal attributes, being able to adapt its forecasting to the current DO concentrations. In contrast, the DNN-Transformer posed a much more rigid approach, tending to maintain the data trend for different DO measurements during the same assessed period, leading to a significant difference between the forecasted and real measured DO. The capacity of the GNN-SAGE to adapt to the input spatiotemporal information from the monitoring stations, reveals its potential to be used as a real-time monitoring tool for water quality forecasting.

5.2. Analysis of Results of the SHAP Analysis

Figure 13 shows the results of the SHAP analysis for the 6-hour forecasting horizon, where the attributes are arranged in a descending manner from the variable with the most influence to the one with the least influence when used to determine future DO concentrations. The rightmost bar is a scale for the feature value based on the correlation between the variable and the forecasted attribute, indicating that an elevated correlation also has an elevated value. SHAP values greater than zero indicate that positive attribute values contribute positively to DO forecasting, while negative values influence the model's prediction negatively.
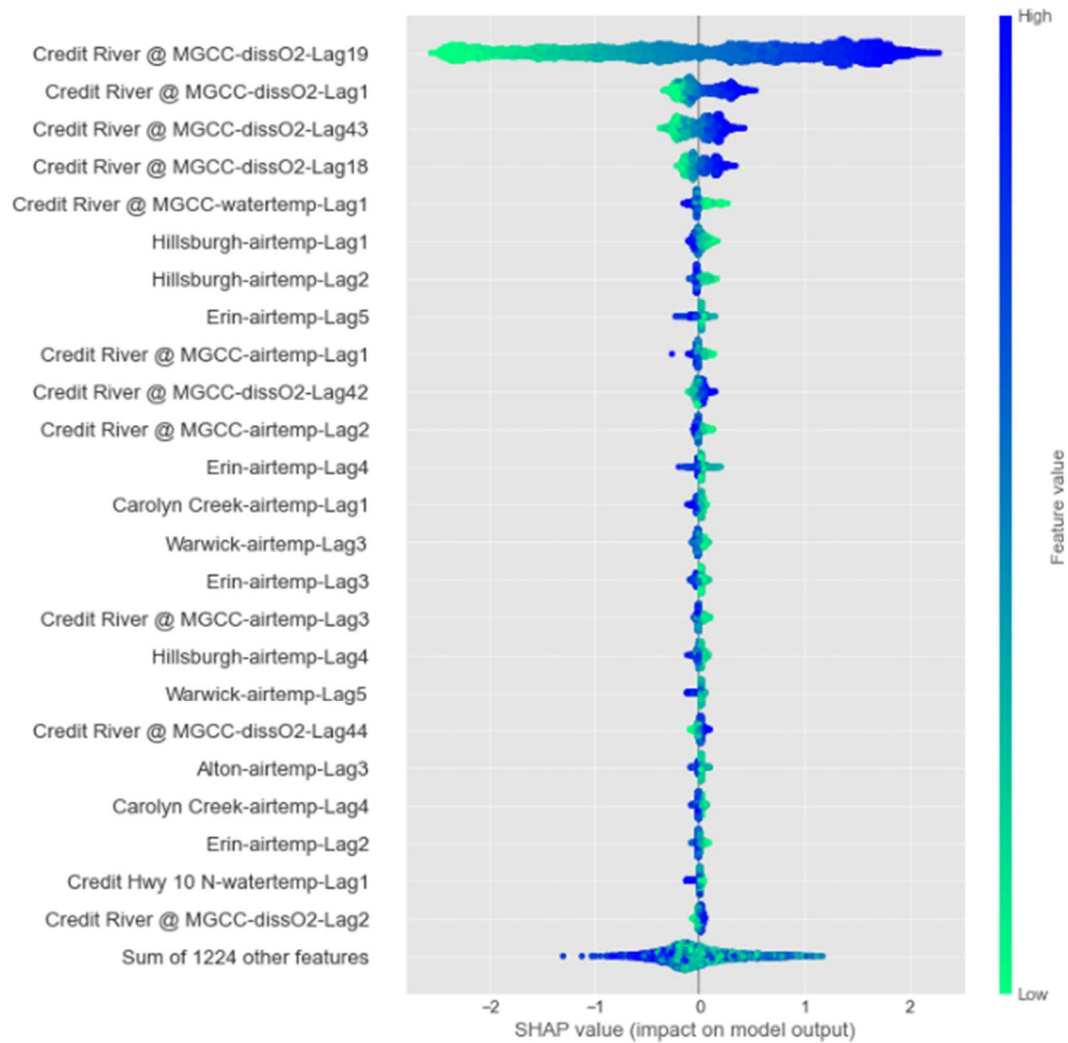
Figure 13. Results from the SHAP analysis for predicting DO concentrations 6 h ahead using GNN-SAGE model.

Figure 13 elucidates that using past DO data, i.e., "dissO2-Lag19," "dissO2-Lag1", "dissO2-Lag43", "dissO2-Lag18", from the reference station "Credit River @ MGCC" provide the most significant amount of spatiotemporal data for the model's forecasting. After that, water and air temperatures are other important factors contributing to the model's DO predictions. For these variables, verifying the relevant influence of spatiotemporal data from the neighbouring stations, such as "Hillsburgh" and "Erin" is important, indicating that the model benefits from their information.

The relevant influence of the temperature attribute over the model's result was expected. Figure 2, depicted the elevated correlation between dissolved oxygen and temperature, once elevated temperatures directly affect the oxygen solubility on the water, also contributing to the photosynthesis process and consequent oxygen concentration on the watershed. The importance of the temperature over future estimations of DO was also studied and verified on previous works [21,33,73], attesting its significant contribution on forecasting models.

The SHAP analysis findings allow better insight on the working of the GNN-SAGE model, showing that it can satisfactorily capture the relationship between the input variables, from the study station and its neighbouring locations. The spatiotemporal data retrieved from the dataset by the model improved its forecasting capability, with relevant influence of past DO and temperature data.

5.3. Analysis of the Coparison Between the GNN-SAGE Results and Literature-Found Values

In order to understand what is place the proposed GNN-SAGE approach among the current forecasting models', the results yielded by the graph-based model in the current study were compared to the ones found in the literature. Nonetheless, such comparison may not provide a truthful picture between the compared models, meaning that their results are inherent to their specific methodologies, input data, geographic location, and application. Yet, a comparison is still a viable approach to determine how well models can perform when compared to their peers. To facilitate the comparison between the proposed GNN-SAGE with the results from the literature, the following Table 1 compiles the results for different metrics for this model when applied to a 6-h forecasting horizon, and Table 2 presents the results from the literature. For further information on the applied metrics, we recommend reference [74].

Table 1. Values of the error metrics for a 6 h forecasting horizon using GNN-SAGE paradigm.

Metric	Value
RMSE	0.34 ppm
nRMSE	3.17%
MAE	0.23 ppm
nMAE	2.14%
MAPE	2.22%
MBE	0.01 ppm
Forecast Skill	74.30%
R^2	97.63%

Table 2. Values for dissolved oxygen concentrations forecasting found in the literature.

Model	Metric Value	Author
Delft3D	RMSE 1.18 mg/L	Oliveira et al. [7]
Delft3D	MAE 1.03 mg/L MAPE 15.9%	Curbani et al. [75]
Prophet	RMSE 0.71 mg/L MAE 0.55 mg/L	Kogekar et al. [76]
LSSVM-BA	RMSE Mean value 0.79 mg/L MAE Mean value 0.94 mg/L	Yaseen et al. [77]
Bi-LSTM	RMSE 0.2 mg/L	Ahmed et al. [24]

	MAE	
	0.15 mg/L	
CEEMDAN–CNN–LSTM	RMSE	Sha et al.
	0.26 mg/L for 4-h forecasting horizon	[26]
	0.28 mg/L for 8-h forecasting horizon	
	0.31 mg/L for 12-h forecasting horizon	
	0.34 mg/L for 16-h forecasting horizon	
	0.39 mg/L for 20-h forecasting horizon	
	0.48 mg/L for 24-h forecasting horizon	
	(Average RMSE of 0.34 mg/L)	
	MAPE	
	2.55% for 4-h forecasting horizon	
	2.79% for 8-h forecasting horizon	
	3.00% for 12-h forecasting horizon	
	3.30% for 16-h forecasting horizon	
	3.65% for 20-h forecasting horizon	
	4.56% for 24-h forecasting horizon	
	(Average MAPE of 3.31%)	

In references [7] and [75], DO was modeled by the physical model of Delft3D for rivers located in Portugal and Brazil, respectively. From the comparison between the GNN-SAGE results and the ones using a physics-based approach, it can be drawn that the graph model provides superior results considering RMSE and MAPE, substantially improving DO forecasting. Also, the MAPE comparison allows us to conclude that the GNN-SAGE provides more accurate estimations for DO than the model used in [75]. Furthermore, it is worth noting that GNN-SAGE, a graph-based model, can inherently process multi-spatiotemporal information contained in the input dataset, without requiring explicit programming and/or feature preprocessing. Another advantage of the GNN-SAGE paradigm is its simpler implementation: differently from the Delft3D model, which models the DO levels by the finite difference method and thus requires the domain to be defined by a mesh grid, GNN-SAGE uses only the inputs presented in Figure 2.

In reference [76], the authors implemented the Prophet ML model. The Prophet was developed and made available by Facebook's Core Data Science group [78], and is a cumulative approach that excels in analyzing and forecasting non-linear data trends [79]. The Prophet was implemented for a DO forecasting task for a river in India in their work. Comparing GNN-SAGE results with ones in [76], the proposed model clearly provides superior results over a simpler ML model implementation such as the Prophet. The superiority of the GNN-SAGE is expected in this scenario, once it can process and identify seasonality and spatial information from the dataset, whilst Prophet may handle temporal data only. From this comparison, we can conclude that spatial data plays a major role in determining future DO levels by incorporating relevant data for the model's prediction, as stated from the SHAP analysis results in Figure 13.

In work [77], the authors proposed a hybrid model called on Least Square Support Vector Machine-Bat Algorithm (LSSVM-BA) for monthly DO estimation in the USA. Their results for the best-performing LSSVM-BA for monthly DO predictions were, on average, 0.79 mg/L and 0.94 mg/L for RMSE and MAE metrics, respectively. Directly comparing these results with GNN-SAGE's, the graph-based model surpassed the literature-found results with a significant difference on both metrics, again proving to be a superior approach to this task. In studies [24] and [26], different DL paradigms were implemented for DO forecasting. Interestingly, among the literature found results

in Table 2, the DL models were the best-performing ones, offering significant improvement for DO estimation over both physical and traditional ML approaches, achieving error values within the same order of magnitude. In work [24], the Bi-LSTM was implemented to model future DO levels in Pakistan using remote sensing data up to 3 time steps ahead. The study [26] used a hybrid model consisting of CNN-LSTM and CEENDAM for predicting DO in China up to 24 hours ahead. The CNN-LSTM hybrid configuration is well known to be able to process spatial and temporal data through its convolutional and recurrent structures, respectively. Both papers reached similar results, for DO forecasting regarding the RMSE metric. However, comparison against GNN-SAGE proved that the proposed paradigm in this work could surpass the results for the DL models by a significant margin, providing more accurate and precise DO forecasting.

6. Conclusions

The present work proposes a cutting-edge forecasting model to determine future DO levels. Using historical data from Credit River, the model was built using spatiotemporal information from monitoring stations and then compared against the XGBoost and the DNN-Transformer models. The best GNN-SAGE configuration was obtained using 72 h time-lag and input parameters Dissolved Oxygen, Air Temperature, Precipitation, and Water Temperature from the reference station "Credit River @ MGCC" and its neighbouring ones.

From the results for the 6-h forecasting horizon, it is possible to notice that the proposed GNN-SAGE can gather influence from the provided spatiotemporal information and be able to identify changes in DO levels. This results in an improvement of 8% considering the RMSE metric when compared with both XGBoost and DNN-Transformer, reaching an R^2 value over 97%. The SHAP analysis results provided further understanding over the model's predicted DO values. Its outcomes elucidated the importance of spatial data coming from the neighboring stations. It also showed the most significant variables: past DO, air, and water temperature.

Comparing GNN-SAGE results with the ones found in the literature, it was possible to attest the superior performance of the graph-based model, which surpassed every assessed model in Table 2 with a significant margin for both RMSE and R^2 values. These findings made the GNN-SAGE model a valuable DO forecasting tool, offering precise and cutting-edge predictions in urban river watersheds.

For future applications of the proposed GNN-SAGE model, other hydrological parameters could be assessed using this approach and directly determining the water quality index. Furthermore, more spatial information could be added to future implementations of the GNN-SAGE, aiming to reduce its geographical bias once it was modeled and validated for only one location. To this end, spatiotemporal data from different rivers could be added to the model, developing its generalization and broadening its geographic application.

Developing a precise and accurate model to forecast real-time hydrological parameters is a relevant topic in the current socio-economic-environmental scenario. Technologies like GNN-SAGE have the potential to enhance the creation of new legislation regarding the protection and sustainable use of watersheds, as well as to support stake-holders in their decision-making and risk management strategies, mitigating further damage from human actions over the aquatic environment when used as a real-time forecasting tool.

Data Availability Statement: The original dataset can be retrieved from https://cvc.ca/real-timemonitoring/ (accessed on 22 September 2023). The algorithms and datasets used can be downloaded from https://drive.google.com/drive/folders/136RH-G-nPVScO7Ln7OOC0WEYl3kk5kDW   and   https://drive.google.com/drive/folders/13Ef-_EklzJze8pZx1oIDoQFKU304d7NF, respectively (accessed on 22 Semptember 2023).

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

References

1. *Paul J. Crutzen and the Anthropocene: A New Epoch in Earth's History*: Benner, S., Lax, G., Crutzen, P.J., Pöschl, U., Lelieveld, J., Brauch, H.G., Eds.; The Anthropocene: Politik—Economics—Society—Science; Springer International Publishing: Cham, 2021; Vol. 1; ISBN 978-3-030-82201-9.

2. Freeman, L.A.; Corbett, D.R.; Fitzgerald, A.M.; Lemley, D.A.; Quigg, A.; Steppe, C.N. Impacts of Urbanization and Development on Estuarine Ecosystems and Water Quality. *Estuaries Coasts* **2019**, *42*, 1821–1838, doi:10.1007/s12237-019-00597-z.

3. Rajkumar, H.; Naik, P.K.; Rishi, M.S. A Comprehensive Water Quality Index Based on Analytical Hierarchy Process. *Ecol. Indic.* **2022**, *145*, 109582, doi:10.1016/j.ecolind.2022.109582.

4. Regier, P.J.; González-Pinzón, R.; Van Horn, D.J.; Reale, J.K.; Nichols, J.; Khandewal, A. Water Quality Impacts of Urban and Non-Urban Arid-Land Runoff on the Rio Grande. *Sci. Total Environ.* **2020**, *729*, 138443, doi:10.1016/j.scitotenv.2020.138443.

5. Giri, S. Water Quality Prospective in Twenty First Century: Status of Water Quality in Major River Basins, Contemporary Strategies and Impediments: A Review. *Environ. Pollut.* **2021**, *271*, 116332, doi:10.1016/j.envpol.2020.116332.

6. Stajkowski, S.; Zeynoddin, M.; Farghaly, H.; Gharabaghi, B.; Bonakdari, H. A Methodology for Forecasting Dissolved Oxygen in Urban Streams. *Water* **2020**, *12*, 2568, doi:10.3390/w12092568.

7. Oliveira, V.H.; Sousa, M.C.; Morgado, F.; Dias, J.M. Modeling the Impact of Extreme River Discharge on the Nutrient Dynamics and Dissolved Oxygen in Two Adjacent Estuaries (Portugal). *J. Mar. Sci. Eng.* **2019**, *7*, 412, doi:10.3390/jmse7110412.

8. Waldron, M.C.; Wiley, J.B. *Water Quality and Processes Affecting Dissolved Oxygen Concentrations in the Blackwater River, Canaan Valley, West Virginia*; U.S. Department of the Interior, US Geological Survey: Charleston, West Virginia, 1996; p. 91;.

9. Yang, F.; Liang, D. Random-Walk Simulation of Non-Conservative Pollutant Transport in Shallow Water Flows. *Environ. Model. Softw.* **2020**, *134*, 104870, doi:10.1016/j.envsoft.2020.104870.

10. 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, doi:10.1021/acs.est.0c06783.

11. Barletta, M.; Lima, A.R.A.; Costa, M.F. Distribution, Sources and Consequences of Nutrients, Persistent Organic Pollutants, Metals and Microplastics in South American Estuaries. *Sci. Total Environ.* **2019**, *651*, 1199–1218, doi:10.1016/j.scitotenv.2018.09.276.

12. Wang, J.; Fu, Z.; Qiao, H.; Liu, F. Assessment of Eutrophication and Water Quality in the Estuarine Area of Lake Wuli, Lake Taihu, China. *Sci. Total Environ.* **2019**, *650*, 1392–1402, doi:10.1016/j.scitotenv.2018.09.137.

13. Varadharajan, C.; Appling, A.P.; Arora, B.; Christianson, D.S.; Hendrix, V.C.; Kumar, V.; Lima, A.R.; Müller, J.; Oliver, S.; Ombadi, M.; et al. Can Machine Learning Accelerate Process Understanding and Decision-Relevant Predictions of River Water Quality? *Hydrol. Process.* **2022**, *36*, e14565, doi:10.1002/hyp.14565.

14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444, doi:10.1038/nature14539.

15. Rocha, P.A.C.; Santos, V.O. Global Horizontal and Direct Normal Solar Irradiance Modeling by the Machine Learning Methods XGBoost and Deep Neural Networks with CNN-LSTM Layers: A Case Study Using the GOES-16 Satellite Imagery. *Int. J. Energy Environ. Eng.* **2022**, *13*, 1271–1286, doi:10.1007/s40095-022-00493-6.

16. Costa Rocha, P.A.; Johnston, S.J.; Oliveira Santos, V.; Aliabadi, A.A.; Thé, J.V.G.; Gharabaghi, B. Deep Neural Network Modeling for CFD Simulations: Benchmarking the Fourier Neural Operator on the Lid-Driven Cavity Case. *Appl. Sci.* **2023**, *13*, 3165, doi:10.3390/app13053165.

17. Marinho, F.P.; Rocha, P.A.C.; Neto, A.R.R.; Bezerra, F.D.V. Short-Term Solar Irradiance Forecasting Using CNN-1D, LSTM, and CNN-LSTM Deep Neural Networks: A Case Study With the Folsom (USA) Dataset. *J. Sol. Energy Eng.* **2023**, *145*, 041002, doi:10.1115/1.4056122.

18. Huang, C.-J.; Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. *Sensors* **2018**, *18*, 2220, doi:10.3390/s18072220.

19. Demolli, H.; Dokuz, A.S.; Ecemis, A.; Gokcek, M. Wind Power Forecasting Based on Daily Wind Speed Data Using Machine Learning Algorithms. *Energy Convers. Manag.* **2019**, *198*, 111823, doi:10.1016/j.enconman.2019.111823.

20. Chimmula, V.K.R.; Zhang, L. Time Series Forecasting of COVID-19 Transmission in Canada Using LSTM Networks. *Chaos Solitons Fractals* **2020**, *135*, 109864, doi:10.1016/j.chaos.2020.109864.

21. Csábrági, A.; Molnár, S.; Tanos, P.; Kovács, J. Application of Artificial Neural Networks to the Forecasting of Dissolved Oxygen Content in the Hungarian Section of the River Danube. *Ecol. Eng.* **2017**, *100*, 63–72, doi:10.1016/j.ecoleng.2016.12.027.

22. Lu, H.; Ma, X. Hybrid Decision Tree-Based Machine Learning Models for Short-Term Water Quality Prediction. *Chemosphere* **2020**, *249*, 126169, doi:10.1016/j.chemosphere.2020.126169.

23. Xu, C.; Chen, X.; Zhang, L. Predicting River Dissolved Oxygen Time Series Based on Stand-Alone Models and Hybrid Wavelet-Based Models. *J. Environ. Manage.* **2021**, *295*, 113085, doi:10.1016/j.jenvman.2021.113085.

24. Ahmed, M.; Mumtaz, R.; Anwar, Z.; Shaukat, A.; Arif, O.; Shafait, F. A Multi–Step Approach for Optically Active and Inactive Water Quality Parameter Estimation Using Deep Learning and Remote Sensing. *Water* **2022**, *14*, 2112, doi:10.3390/w14132112.

25. 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, doi:10.1007/s00477-020-01776-2.

26. Sha, J.; Li, X.; Zhang, M.; Wang, Z.-L. Comparison of Forecasting Models for Real-Time Monitoring of Water Quality Parameters Based on Hybrid Deep Learning Neural Networks. *Water* **2021**, *13*, 1547, doi:10.3390/w13111547.

27. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. Spatiotemporal Analysis of Bidimensional Wind Speed Forecasting: Development and Thorough Assessment of LSTM and Ensemble Graph Neural Networks on the Dutch Database. *Energy* **2023**, *278*, 127852, doi:10.1016/j.energy.2023.127852.

28. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. A New Graph-Based Deep Learning Model to Predict Flooding with Validation on a Case Study on the Humber River. *Water* **2023**, *15*, 1827, doi:10.3390/w15101827.

29. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. Spatiotemporal Air Pollution Forecasting in Houston-TX: A Case Study for Ozone Using Deep Graph Neural Networks. *Atmosphere* **2023**, *14*, 308, doi:10.3390/atmos14020308.

30. Li, D.; Zhao, W.; Hu, J.; Zhao, S.; Liu, S. A Long-Term Water Quality Prediction Model for Marine Ranch Based on Time-Graph Convolutional Neural Network. *Ecol. Indic.* **2023**, *154*, 110782, doi:10.1016/j.ecolind.2023.110782.

31. Liu, G.; Jiang, Y.; Zhong, K.; Yang, Y.; Wang, Y. A Time Series Model Adapted to Multiple Environments for Recirculating Aquaculture Systems. *Aquaculture* **2023**, *567*, 739284, doi:10.1016/j.aquaculture.2023.739284.

32. Ni, Q.; Cao, X.; Tan, C.; Peng, W.; Kang, X. An Improved Graph Convolutional Network with Feature and Temporal Attention for Multivariate Water Quality Prediction. *Environ. Sci. Pollut. Res.* **2022**, *30*, 11516–11529, doi:10.1007/s11356-022-22719-0.

33. Fang, Y.; Liu, H. A Spatiotemporal Dissolved Oxygen Prediction Model Based on Graph Attention Networks Suitable for Missing Data. *Environ. Sci. Pollut. Res.* **2023**, *30*, 82818–82833, doi:10.1007/s11356-023-28030-w.

34. Schuster-Wallace, C.J.; Murray, S.J.; McBean, E.A. Integrating Social Dimensions into Flood Cost Forecasting. *Water Resour. Manag.* **2018**, *32*, 3175–3187, doi:10.1007/s11269-018-1983-8.

35. Allen, B.; Mandrak, N.E. Historical Changes in the Fish Communities of the Credit River Watershed. *Aquat. Ecosyst. Health Manag.* **2019**, *22*, 316–328, doi:10.1080/14634988.2019.1672463.

36. McGovarin, S.; Nishikawa, J.; Metcalfe, C.D. Vitellogenin Induction in Mucus from Brook Trout (Salvelinus Fontinalis). *Bull. Environ. Contam. Toxicol.* **2022**, *108*, 878–883, doi:10.1007/s00128-021-03440-z.

37. Champagne, O.; Arain, M.A.; Leduc, M.; Coulibaly, P.; McKenzie, S. Future Shift in Winter Streamflow Modulated by the Internal Variability of Climate in Southern Ontario. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 3077–3096, doi:10.5194/hess-24-3077-2020.

38. Rincón, D.; Khan, U.; Armenakis, C. Flood Risk Mapping Using GIS and Multi-Criteria Analysis: A Greater Toronto Area Case Study. *Geosciences* **2018**, *8*, 275, doi:10.3390/geosciences8080275.

39. Rincón, D.; Velandia, J.F.; Tsanis, I.; Khan, U.T. Stochastic Flood Risk Assessment under Climate Change Scenarios for Toronto, Canada Using CAPRA. *Water* **2022**, *14*, 227, doi:10.3390/w14020227.

40. Chu, C.; Minns, C.K.; Lester, N.P.; Mandrak, N.E. An Updated Assessment of Human Activities, the Environment, and Freshwater Fish Biodiversity in Canada. *Can. J. Fish. Aquat. Sci.* **2015**, *72*, 135–148, doi:10.1139/cjfas-2013-0609.

41. Jane, S.F.; Hansen, G.J.A.; Kraemer, B.M.; Leavitt, P.R.; Mincer, J.L.; North, R.L.; Pilla, R.M.; Stetler, J.T.; Williamson, C.E.; Woolway, R.I.; et al. Widespread Deoxygenation of Temperate Lakes. *Nature* **2021**, *594*, 66–70, doi:10.1038/s41586-021-03550-y.

42. Roman, M.R.; Brandt, S.B.; Houde, E.D.; Pierson, J.J. Interactive Effects of Hypoxia and Temperature on Coastal Pelagic Zooplankton and Fish. *Front. Mar. Sci.* **2019**, *6*, 139, doi:10.3389/fmars.2019.00139.

43. Stajkowski, S.; Hotson, E.; Zorica, M.; Farghaly, H.; Bonakdari, H.; McBean, E.; Gharabaghi, B. Modeling Stormwater Management Pond Thermal Impacts during Storm Events. *J. Hydrol.* **2023**, *620*, 129413, doi:10.1016/j.jhydrol.2023.129413.

44. Yang, R.; Sun, H.; Chen, B.; Yang, M.; Zeng, Q.; Zeng, C.; Huang, J.; Luo, H.; Lin, D. Temporal Variations in Riverine Hydrochemistry and Estimation of the Carbon Sink Produced by Coupled Carbonate Weathering with Aquatic Photosynthesis on Land: An Example from the Xijiang River, a Large Subtropical Karst-Dominated River in China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 13142–13154, doi:10.1007/s11356-020-07872-8.

45. Xuan, Y.; Tang, C.; Liu, G.; Cao, Y. Carbon and Nitrogen Isotopic Records of Effects of Urbanization and Hydrology on Particulate and Sedimentary Organic Matter in the Highly Urbanized Pearl River Delta, China. *J. Hydrol.* **2020**, *591*, 125565, doi:10.1016/j.jhydrol.2020.125565.

46. Dawoud, I.; Abonazel, M.R. Robust Dawoud–Kibria Estimator for Handling Multicollinearity and Outliers in the Linear Regression Model. *J. Stat. Comput. Simul.* **2021**, *91*, 3678–3692, doi:10.1080/00949655.2021.1945063.

47. Chan, J.Y.-L.; Leow, S.M.H.; Bea, K.T.; Cheng, W.K.; Phoong, S.W.; Hong, Z.-W.; Chen, Y.-L. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics* **2022**, *10*, 1283, doi:10.3390/math10081283.

48. Oliveira Santos, V.; Costa Rocha, P.A.; Gharabaghi, B.; Thé, J.V.G. Graph-Based Deep Learning Model for Forecasting Chloride Concentration in Urban Streams to Protect Salt-Vulnerable Areas. **2023**, doi:10.20944/preprints202308.0450.v1.

49. Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. *Energies* **2020**, *13*, 3764, doi:10.3390/en13153764.

50. Trebing, K.; Mehrkanoon, S. Wind Speed Prediction Using Multidimensional Convolutional Neural Networks. In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI); December 2020; pp. 713–720.

51. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: San Francisco California USA, August 13 2016; pp. 785–794.

52. Ghojogh, B.; Crowley, M. The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial 2023.

53. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967, doi:10.1007/s10462-020-09896-5.

54. Bondi, A.B. Characteristics of Scalability and Their Impact on Performance. In Proceedings of the Proceedings of the second international workshop on Software and performance   - WOSP '00; ACM Press: Ottawa, Ontario, Canada, 2000; pp. 195–203.

55. Wang, Y.; Pan, Z.; Zheng, J.; Qian, L.; Li, M. A Hybrid Ensemble Method for Pulsar Candidate Classification. *Astrophys. Space Sci.* **2019**, *364*, 139, doi:10.1007/s10509-019-3602-4.

56. Zhang, W.; Wu, C.; Zhong, H.; Li, Y.; Wang, L. Prediction of Undrained Shear Strength Using Extreme Gradient Boosting and Random Forest Based on Bayesian Optimization. *Geosci. Front.* **2021**, *12*, 469–477, doi:10.1016/j.gsf.2020.03.007.

57. Parsa, A.B.; Movahedi, A.; Taghipour, H.; Derrible, S.; Mohammadian, A. (Kouros) Toward Safer Highways, Application of XGBoost and SHAP for Real-Time Accident Detection and Feature Analysis. *Accid. Anal. Prev.* **2020**, *136*, 105405, doi:10.1016/j.aap.2019.105405.

58. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks.; Vancouver, BC, Canada, February 4 2018.

59. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need 2017.

60. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

61. Chollet, F. *Deep Learning with Python, Second Edition*; Simon and Schuster, 2021; ISBN 978-1-63835-009-5.

62. Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph Convolutional Networks: A Comprehensive Review. *Comput. Soc. Netw.* **2019**, *6*, 11, doi:10.1186/s40649-019-0069-y.

63. Wilson, T.; Tan, P.-N.; Luo, L. A Low Rank Weighted Graph Convolutional Approach to Weather Prediction. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM); November 2018; pp. 627–636.

64. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive Representation Learning on Large Graphs. In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc., 2017; Vol. 30.

65. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24, doi:10.1109/TNNLS.2020.2978386.

66. Labonne, M. Hands-On Graph Neural Networks Using Python.
67. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Stanford* **2013**.
68. Liew, S.S.; Khalil-Hani, M.; Bakhteri, R. Bounded Activation Functions for Enhanced Training Stability of Deep Neural Networks on Visual Pattern Recognition Problems. *Neurocomputing* **2016**, *216*, 718–734, doi:10.1016/j.neucom.2016.08.037.
69. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215, doi:10.1038/s42256-019-0048-x.
70. McGovern, A.; Lagerquist, R.; John Gagne, D.; Jergensen, G.E.; Elmore, K.L.; Homeyer, C.R.; Smith, T. Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 2175–2199, doi:10.1175/BAMS-D-18-0195.1.
71. Gu, Z.; Liu, Y.; Hughes, D.J.; Ye, J.; Hou, X. A Parametric Study of Adhesive Bonded Joints with Composite Material Using Black-Box and Grey-Box Machine Learning Methods: Deep Neuron Networks and Genetic Programming. *Compos. Part B Eng.* **2021**, *217*, 108894, doi:10.1016/j.compositesb.2021.108894.
72. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc., 2017; Vol. 30.
73. Ziyad Sami, B.F.; Latif, S.D.; Ahmed, A.N.; Chow, M.F.; Murti, M.A.; Suhendi, A.; Ziyad Sami, B.H.; Wong, J.K.; Birima, A.H.; El-Shafie, A. Machine Learning Algorithm as a Sustainable Tool for Dissolved Oxygen Prediction: A Case Study of Feitsui Reservoir, Taiwan. *Sci. Rep.* **2022**, *12*, 3649, doi:10.1038/s41598-022-06969-z.
74. Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.C.; Coimbra, C.F.M. History and Trends in Solar Irradiance and PV Power Forecasting: A Preliminary Assessment and Review Using Text Mining. *Sol. Energy* **2018**, *168*, 60–101, doi:10.1016/j.solener.2017.11.023.
75. Curbani, F.E.; Lacerda, K.C.; Curbani, F.; Barreto, F.T.C.; Tadokoro, C.E.; Chacaltana, J.T.A. Numerical Study of Physical and Biogeochemical Processes Controlling Dissolved Oxygen in an Urbanized Subtropical Estuary: Vitória Island Estuarine System, Brazil. *Environ. Model. Assess.* **2022**, *27*, 233–249, doi:10.1007/s10666-021-09787-1.
76. Kogekar, A.P.; Nayak, R.; Pati, U.C. Forecasting of Water Quality for the River Ganga Using Univariate Time-Series Models. In Proceedings of the 2021 8th International Conference on Smart Computing and Communications (ICSCC); July 2021; pp. 52–57.
77. Yaseen, Z.; Ehteram, M.; Sharafati, A.; Shahid, S.; Al-Ansari, N.; El-Shafie, A. The Integration of Nature-Inspired Algorithms with Least Square Support Vector Regression Models: Application to Modeling River Dissolved Oxygen Concentration. *Water* **2018**, *10*, 1124, doi:10.3390/w10091124.
78. Facebook's Core Data Science team Forecasting at Scale Available online: http://facebook.github.io/prophet/ (accessed on 20 September 2023).
79. Taylor, S.J.; Letham, B. *Forecasting at Scale*; PeerJ Preprints, 2017;