ADVANCING EARTH AND SPACE SCIENCES

Earth and Space Science

RESEARCH ARTICLE
10.1029/2023EA003473

Multi-Model Machine Learning Approach Accurately Predicts Lake Dissolved Oxygen With Multiple Environmental Inputs

Shuqi Lin[1,2], **Donald C. Pierson**[2], **Robert Ladwig**[3], **Benjamin M. Kraemer**[4], and **Fenjuan R. S. Hu**[5]

[1]Environment and Climate Change Canada, Canada Centre for Inland Waters, Burlington, ON, Canada, [2]Erken Laboratory and Limnology Department, Uppsala University, Uppsala, Sweden, [3]Center for Limnology, University of Wisconsin-Madison, Madison, WI, USA, [4]IGB Leibniz Institute for Freshwater Ecology and Inland Fisheries, Berlin, Germany, [5]Research Center for Building, Energy, Water and Climate, VIA University College, Horsens, Denmark

Key Points:

- A 2-step machine learning workflow combining Gradient Boost Regressor and long-short-term-memory was applied to simulate dissolved oxygen
- A one-dimensional process-based hydrodynamic model provides ML models with indices of lake thermal structure and mixing
- In a polymictic lake, the 2-step mixed machine learning workflow showed over 90% true positive rate of hypolimnetic hypoxia detection

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Lin,
Shuqi.lin@ec.gc.ca

Citation:

Lin, S., Pierson, D. C., Ladwig, R., Kraemer, B. M., & Hu, F. R. S. (2024). Multi-model machine learning approach accurately predicts lake dissolved oxygen with multiple environmental inputs. *Earth and Space Science*, *11*, e2023EA003473. https://doi.org/10.1029/2023EA003473

Received 21 DEC 2023
Accepted 13 MAY 2024

Author Contributions:

Conceptualization: Shuqi Lin, Donald C. Pierson
Data curation: Shuqi Lin, Robert Ladwig, Benjamin M. Kraemer, Fenjuan R. S. Hu
Formal analysis: Shuqi Lin
Funding acquisition: Donald C. Pierson
Investigation: Shuqi Lin
Methodology: Shuqi Lin
Project administration: Donald C. Pierson
Software: Shuqi Lin, Robert Ladwig
Supervision: Donald C. Pierson
Validation: Shuqi Lin
Visualization: Shuqi Lin
Writing – original draft: Shuqi Lin

Abstract As a key water quality parameter, dissolved oxygen (DO) concentration, and particularly changes in bottom water DO is fundamental for understanding the biogeochemical processes in lake ecosystems. Based on two machine learning (ML) models, Gradient Boost Regressor (GBR) and long-short-term-memory (LSTM) network, this study developed three ML model approaches: direct GBR; direct LSTM; and a 2-step mixed ML model workflow combining both GBR and LSTM. They were used to simulate multi-year surface and bottom DO concentrations in five lakes. All approaches were trained with readily available environmental data as predictors. Indices of lake thermal structure and mixing provided by a one-dimensional (1-D) hydrodynamic model were also included as predictors in the ML models. The advantages of each ML approach were not consistent for all the tested lakes, but the best one of them was defined that can estimate DO concentration with coefficient of determination (R^2) up to 0.6–0.7 in each lake. All three approaches have normalized mean absolute error (NMAE) under 0.15. In a polymictic lake, the 2-step mixed model workflow showed better representation of bottom DO concentrations, with a highest true positive rate (TPR) of hypolimnetic hypoxia detection of over 90%, while the other workflows resulted in, TPRs are around 50%. In most of the tested lakes, the predicted surface DO concentrations and variables indicating stratified conditions (i.e., Wedderburn number and the temperature difference between surface and bottom water) are essential for simulating bottom DO. The ML approaches showed promising results and could be used to support short- and long-term water management plans.

Plain Language Summary Dissolved oxygen (DO) concentrations is the essential water quality parameter in lake systems. Nowadays, with the development of data-driven machine learning (ML) models, prediction of DO concentrations can be achieved via these models in lakes with long-term DO concentration observations. This study developed three ML model approaches with one mixed two kind of ML models, and test them in five lakes. Readily available environmental data and the derived hydrodynamic data from process-based hydrodynamic model were used as predictors. All three ML approaches showed promising results, and the mixed ML approach show better skill in the lake stratifying and mixing irregularly. To predict hypoxia in the bottom of the lake, the surface DO concentrations and variables indicating water column stratification are important.

1. Introduction

Dissolved oxygen (DO) is an essential ecosystem variable regularly used to assess water quality, responding to changes in phytoplankton photosynthesis, ecosystem respiration, mineralization, and lake mixing. Changes in the duration and frequency of hypolimnetic hypoxia are often used as an indicator of aquatic ecosystem health (Jane et al., 2021). A decline of DO can have a significant impact on water quality by potentially increasing internal nutrient loading and further modifying lake trophic state (Orihel et al., 2017), promoting harmful algal bloom formations (Paerl & Paul, 2012), and reducing fish habitat which eventually can cause severe fish kills (Rao et al., 2014). Hypoxia is becoming more common in the hypolimnion of lakes due to intensifying thermal stratification and loss of water clarity (Jane et al., 2021; North et al., 2014). Variables that regulate DO concentration include atmospheric forcing, convective mixing, algal biomass, nutrient loading, sediment resuspension, and sediment oxygen demand (Müller et al., 2012; Charlton & Lean, 1987; Ladwig et al., 2021a, 2021b).

Writing – review & editing: Shuqi Lin, Donald C. Pierson, Robert Ladwig, Benjamin M. Kraemer, Fenjuan R. S. Hu

A variety of modeling tools, including analytical models (Bouffard et al., 2013; Cortés et al., 2021), and numerical models, that is, coupled hydrodynamic-water-quality models (e.g., Ladwig et al., 2022; Léon et al., 2011) have been developed to simulate lake DO dynamics via parameterizing and simulation of the regulatory processes mentioned above. However, water quality simulations tend to be restricted to an individual lake due to the differing importance of biogeochemical processes within each lake, which must be accounted for by model parameterization.

State-of-art data-driven machine learning (ML) models have been applied in the wide range of water resource research, simulating lake water temperature (Read et al., 2019; Yousefi & Toffolon, 2022) and water quality parameters, for example, phosphorus (P) (Hanson et al., 2020), algal chlorophyll concentrations (Chl) (Kakouei et al., 2021; Lin et al., 2023), and DO (Ziyad Sami et al., 2022). Inspired by the robust performance of ML in capturing the nonlinearity patterns in the systems, multiple ML approaches have been applied to simulate DO concentration, including artificial neural network (ANN), Support Vector Machine (SVM), Extreme Learning Machine (ELM), etc (Dehghani et al., 2022; Zhu & Heddam, 2020; Ziyad Sami et al., 2022). But most of previous research assessed the prediction skill based on single-use of ML model, or compared the skills of multiple ML models in an individual surface water system. The adaptivity of ML models in lakes with various mixing dynamics has not been fully assessed.

In this study, we developed three ML modeling approaches based on two ML models, Gradient Boosting Regressor (GBR) and long-short-term-memory (LSTM) network to simulate multi-year seasonal-scale surface and bottom DO concentrations in five lakes with various sizes and trophic levels. In addition to applying GBR and LSTM directly, we designed a 2-step mixed model workflow by inputting the model residuals from a GBR model into LSTM as the response variable. These ML approaches were trained with available meteorological forcing data. Further, we used a one-dimensional (1-D) hydrodynamic model forced with the same meteorological and hydrological data to provide additional information on lake thermal structure, and ice cover that was additionally included as ML model training features. Transferring knowledge from simulations produced by process-based (PB) models could improve the generalizable pattern learning of ML models (Jia et al., 2021; Read et al., 2019). This hybrid approach has achieved promising results in algal bloom predictions in a mesotrophic lake (Lin et al., 2023), and may also improve predictions of lake DO and, in particular, hypoxia, which is strongly dependent on lake hydrodynamics.

In addition to evaluate these ML approaches in simulating the variability of DO and detecting hypolimnetic hypoxia in the lakes, this study also aims to explore the significant factors regulating DO concentrations in each individual lake. In the following sections, comparison with process-based (PB) models, as well as the limitations, and future applications of the ML approaches in the water management are presented and discussed.

2. Materials and Methods

2.1. Study Sites

This study used data from five lakes: Lake Erken (Sweden), Müggelsee (Germany), Lake Furesø (Denmark), Lake Mendota (USA), and Lake Ekoln (Sweden). Each lake's characteristics are described in Table 1. The detailed monitoring programs in each lake can be found in Supporting Information S1 (Text S1).

DO sampling interval varies among lakes. Since 2007, a multi-parameter YSI profiling system was installed at the Müggelsee observation station providing hourly DO concentration measurements. And since 2015, Lake Erken improved its automated monitoring program to include a YSI profiling system that collects hourly profiles of DO concentrations. For these two lakes, the hourly surface DO concentrations were averaged to provide daily surface values, while the daily the minimum bottom DO concentrations were used to represent the daily bottom values. In Furesø, Lake Mendota, and Ekoln, DO concentrations were recorded by water samples which have biweekly to monthly intervals (Text S1 in Supporting Information S1).

In Furesø, a major restoration project started in 2003 to control the internal loading of phosphorus from the sediment during stratification. Mitigation measures included a combination of hypolimnetic aeration and biomanipulation (Gurkan et al., 2006). Since then, hypoxia has been reduced in the bottom waters of the lake (Johansson et al., 2021).

Table 1
Physical Characteristics of the Lakes

Characteristics	Lake Erken[a]	Müggelsee[b]	Furesø[c]	Lake Mendota[d, e]	Lake Ekoln[f]
Lake area (km^2)	23.7	7.4	9.4	39.6	29.8
Mean/Max depth (m)	9/21	4.9/8	7.4/37.7	12.7/25	15.4/50
Residence time (Years)	7	0.12–0.15	10	4.3	<1
Lake mixing type	Dimictic	Polymictic[g]	Dimictic	Dimictic[d]	Dimictic
Trophic state	Mesotrophic	Eutrophic	Mesotrophic	Eutrophic	Eutrophic
Averaged DO sampling interval during ice-free period (Days)	8 (2004–2014) 1 (2015–2020)	1	18	16	31
Data span (Total years)	2004–2020 (17 years)	2004–2020 (17 years)	1990–2017 (28 years)	1999–2015 (17 years)	1987–2019 (33 years)
Training period (Percentage of data)	2004–2016 (76%)	2004–2016 (76%)	1990–2009 (71%)	1999–2009 (65%)	1987–2008 (67%)
Testing period (Percentage of data)	2017–2020 (24%)	2017–2020 (24%)	2010–2017 (29%)	2010–2015 (35%)	2009–2019 (33%)

[a]Pierson et al., 1992. [b]Kakouei et al., 2022. [c]Gurkan et al., 2006. [d]Farrell et al., 2020. [e]Bennett et al., 1999. [f]Goedkoop et al., 2011. [g]Shatwell and Köhler, 2019.

The invasive aquatic plant species Nuttall's waterweed (*Elodea nuttallii*) was first detected in 2011 in Müggelsee, and has spread rapidly, becoming the most abundant macrophyte species by 2017. The abundance of another invasive species, the dreissenid mussel, increased with the increasing invasive waterweed, following the invasion meltdown hypothesis (Wegner et al., 2019). *E. nuttallii* can largely increase the oxygen production via photosynthesis but also can result in extremely low DO in the bottom water of lakes, developing night-time anoxic conditions (Vilas et al., 2017).

In Lake Mendota, the grazing activity of freshwater zooplankton, that is, Daphnia, leads to a reduction in algal biomass in late spring-early summer, during the so-called clear water phase (Carpenter & Kitchell, 1988). However, the lake also experienced the invasion from the spiny water flea, *Bythotrephes longimanus* (hereafter *Bythotrephes*), during the fall of 2009, which led a decline of the daphnia population, reduced grazing of the spring diatom population, and an overall decline in water clarity. However, although one of the major *Daphnia* species (i.e., *Daphnia pulicaria*) is a preferred prey of *Bythotrephes*, another smaller-bodied Daphnia Mendotae, now increases in spring following the invasion of *Bythotrephes* (see Figure S1 in Supporting Information S1). The combined result of these changes was reduced grazing on spring diatom and probably accelerating organic matter mineralization and hypolimnetic oxygen depletion before summer stratification (Ladwig et al., 2021a, 2021b; Matsuzaki et al., 2021; Rohwer et al., 2023; Walsh et al., 2017). Surface nutrient loadings for Lake Mendota were derived from a calibrated catchment model, PIHM-Lake (see Ladwig et al., 2021a, 2021b for more information).

2.2. Models

2.2.1. Process-Based (PB) Hydrodynamic Models

One-dimensional PB hydrodynamic models were used to estimate metrics that describe lake thermal structure and mixing, and which could also serve as training inputs to the ML models. The 1-D hydrodynamic model, GOTM (General Ocean Turbulence Model (Burchard et al., 1999)); was applied in Lake Erken (Mesman et al., 2022; Moras et al., 2019), and Müggelsee, Furesø, and Lake Ekoln, while GLM (General Lake Model (Hipsey et al., 2019) was applied in Lake Mendota (Ladwig et al., 2021a, 2021b)). The meteorological variables (i.e., air temperature, air pressure, solar radiation, cloud cover, wind speed, precipitation, relative humidity) and river discharge used to train the ML models were also the inputs of the hydrodynamic models (Figure 1). The PB modeled water temperatures in Müggelsee, Furesø and Lake Ekoln were calibrated against temperature profile observations, and the root-mean-squared-error (RMSE) are 1.33°C, 1.29°C, 1.51°C. Ladwig et al., 2021a, 2021b reported GLM was able to simulate water temperature in Lake Mendota with around 1.3°C RMSE and Moras et al. (2019) showed the RMSE of GOTM modeled water temperature in Lake Erken was around 1.1°C. We used the daily vertical profiles of simulated water temperature and eddy diffusion (K_z) obtained from the PB models to derive daily features to train the ML models. The temperature difference (delT) between surface water (averaged over the upper 3 m) and bottom water (bottom layers) was calculated based on modeled temperature profiles.

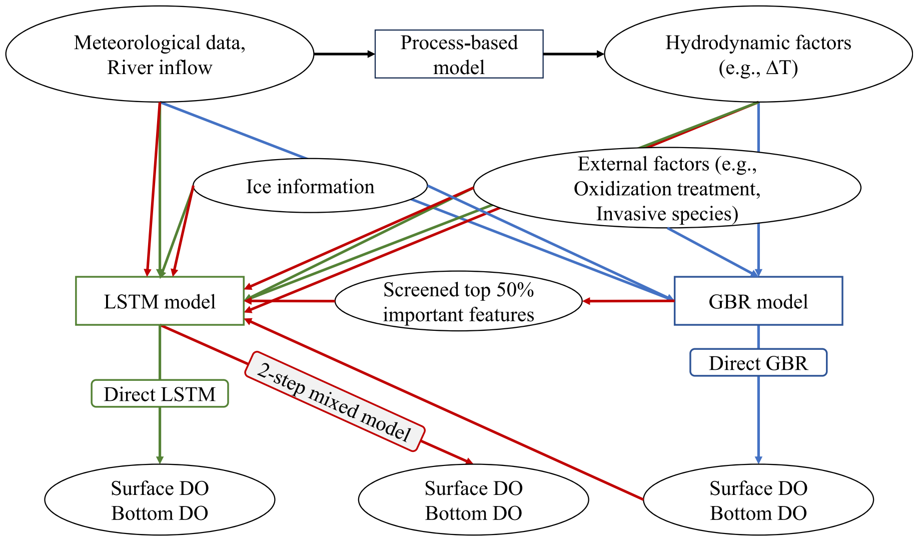

Figure 1. Workflow of three machine learning models. Green arrows represent direct LSTM model, blue arrows represent direct GBR model, and red arrows represent 2-step mixed model.

Mixed layer depth (MLD) was defined as the first depth, from the lake surface, where K_z fell below 5×10^{-5} m^2s^{-1} threshold (Wilson et al., 2020), and Wedderburn number (W_n) was computed based on MLD calculations (Thompson & Imberger, 1980). Here, *Wn* indicates the magnitude of wind-induced upwelling. We used Lake Analyzer (Read et al., 2011) to estimate thermocline depth (thermD) and Schmidt stability (St). These parameters, based on the daily temperature profiles, indicate the extent of mixing, hypolimnetic thickness and the intensity of stratification (Wetzel, 2001), which can be further related to the variability of bottom DO concentrations (Cortés et al., 2021; Foley et al., 2012; North et al., 2014).

2.2.2. Direct LSTM and GBR Models

This study applied two ML models, LSTM and GBR, built by the Scikit-Learn (https://scikit-learn.org/stable/, last access: September 2022) and TensorFlow (https://www.tensorflow.org/, last access: September 2022) libraries in Python.

GBR is a type of tree model, a class of ML models that are most commonly applied in water resource studies, including DO prediction (Heddam & Kisi, 2018; Kisi et al., 2020). The model iteratively generates an ensemble of estimator trees with each tree improving upon the performance of the previous one (Friedman, 2001). The hyperparameters, including *n_estimators, max_depth, learning_rate*, *subsample*, in GBR are optimized via Randomized Search (*RandomizedSearchCV* function within Scikit-Learn library) based on 5-fold cross validation. The loss function used within the model was “huber”, a combination of the squared error and absolute error of simulation. Note that the hyperparameters may differ for each lake, the model was designed to go through hyperparameter tunning when training data changed. The GBR model can rank the feature importance for each predictive target (Friedman, 2001), illustrating the key factors which regulate the DO concentrations.

LSTM is a recurrent neural network, built for sequential and time-series modeling (Hochreiter & Schmidhuber, 1997). This model architecture has also been applied in many water resource studies (Read et al., 2019), and has achieved promising results in predicting harmful algal blooms in Lake Erken (Lin et al., 2023). We built a LSTM with three hidden layers each with 50 neurons in every layer, and each of them is followed by a dropout layer with 0.2 dropout rate for regularization. The numbers of batchs and epochs are set as 20–100, respectively. The hyperparameters were chosen based on the tradeoff between computational cost and model performance. The data were scaled to a given range for generalization purposes via “MinMaxScaler” function, and “Mean Absolute Error” was used as loss function. The time step of LSTM was set to 7 days, which means the memory of all the training features within the previous 7 days was used to train the model and predict the targets.

The direct applications of these two models involved using the training features described below (Table 2) and corresponding targets (Surface and bottom DO concentrations) along the timeseries in the training periods to train

Table 2
Training Features in Each Lake

Features (data types)	Lake Erken	Müggelsee	Furesø	Lake mendota	Lake Ekoln
River discharge (Metrics)	■	■	■	■	■
Air temperature (Metrics)	■	■	■	■	■
Air pressure (Metrics)		■	■		■
Precipitation (Metrics)	■	■		■	■
Wind speed (Metrics)	■	■	■	■	■
Humidity (Metrics, 0–100)	■	■	■		■
Shortwave radiation (Metrics)	■	■		■	■
Cloud cover (Metrics, 0–1)	■	■	■		■
delT (Metrics)	■	■	■	■	■
Accumulated bottom water temperature over 10 days (Metrics)	■				
Ice duration (Category)[a]	■	■		■	■
Days from ice-off date (Category)[b]	■	■		■	■
MLD (Metrics)	■	■	■	■	■
W_n (Metrics)	■	■	■	■	■
thermD (Metrics)	■	■	■		■
Water treatment (Binary)			■		
Invasive species (Binary)		■		■	
Daphnia (Category)[c]				■	
Accumulated phosphate from river loading (Metrics)				■	
Dissolved organic nutrients from river loading (Metrics)				■	
Inflow temperature (Metrics)				■	

[a]4 levels: ice duration over 60 days, 30–60 days, less than 30 days, and no ice duration. [b]7 levels: over 30 days before ice-off date, 30–20 days before ice-off date, 10–20 days before ice-off date, 10 days before or 10 days after ice-off date, 10–20 days after ice-off date, 20–30 days after ice-off date, over 30 days after ice-off date. [c]4 levels: Daphnia biomass >400/m^3, 200–400/m^3, 50–200/m^3, <50/m^3.

the models. Model performance (validation) was tested by inputting the features along the timeseries in the testing periods and comparing the predictive targets with the measurements (Figure 1).

2.2.3. 2-Step Mixed ML Model Workflow

The 2-step mixed ML model workflow integrates both the GBR and the LSTM model. First, the GBR was used to simulate both surface and bottom DO concentrations, and to rank the importance of training features (Friedman, 2001) affecting surface and bottom DO, respectively. Secondly, after ranking by importance only the features in the top 50% were retained in the training of LSTM models so that the more computationally demanding LSTM training was accelerated by only considering the most significant features. Also, the predictive daily values of DO concentration from GBR were added into the training data set of LSTM as training features (Figure 1). Based on the results from the direct GBR model, the seasonal variability of DO concentrations could be represented, so that the GBR results are reasonable first estimates and can be used as initial values for the training processes of LSTM models.

2.3. Training Features

The general training features used in every tested lake are daily meteorological data, river inflow data, ice information, and derived hydrodynamic factors. However, given that the physical and biogeochemical characteristics and data availability varied for each lake, each lake has its unique set of training features (Table 2).

In addition to the physical variables used to drive PB hydrodynamic models, daily hydrodynamic variables derived from PB models (i.e., delT, MLD, thermD, W_n) were involved in training features. In Lake Erken, training features also included accumulated bottom water temperature over past 10 days calculated from observation.

The duration of ice cover period could affect the frequency, intensity, and occurrence of winter mixing events, and further impact the renewal of deep water. Ice related features, that is, ice duration and days from ice-off date, were included as training features for the lakes that were routinely ice covered (Lake Erken, Müggelsee, Lake Mendota, Lake Ekoln; Table 2). Note that we aimed to avoid the direct usage of time-related features, since the seasonality related to time strongly affects DO and tends to overwhelm the impact from other features. Also, tree models tend to perform better when handling categorical data. Thus, the ice duration feature was converted into 4-level categorical feature according to the length of ice duration in the previous winter. The days from ice-off date feature was also converted into 7-level categorical feature according to the days before or after ice-off date (Table 2).

In Lake Mendota, zooplankton, that is, Daphnia, density which were recorded biweekly were interpolated linearly into daily values. Original data shows the biomass of Daphnia per m^3, but since the data were sparse, we categorized Daphnia into 4 levels (Table 2). Accumulated Phosphate and Dissolved organic nutrients were estimated by regression models using discharge and nutrient concentration data from USGS gages (see data in Ladwig et al., 2021a, 2021b).

To consider other external factors in the specific lakes, we added external training factors, to account for invasive species in Müggelsee, and Lake Mendota, and hypolimnetic aeration in Furesø. These factors were set as binary numbers, with one representing period after the invasion in Müggelsee and Lake Mendota, or on-going water treatment operation in Furesø.

In addition, DO in the surface and bottom waters were predicted sequentially by the ML models, with predictive surface DO being included in the training features of the bottom DO.

2.4. Model Evaluation

Mean absolute error (MAE), Root-mean-squared-error (RMSE) and correlation coefficient (R^2) of the modeled surface and bottom DO concentration were used to evaluate each model approach in each individual lake, respectively. Since the models were tested in the lakes with various trophic states, the MAE were normalized by the range of observed concentrations in order to be used to conduct inter-lake comparison (Equation 1).

$$\text{NMAE} = \text{MAE}/(\max(y) - \min(y)) \tag{1}$$

To assess the uncertainty induced by variations in the training data, we randomly removed tow individual years data (6%–15% of data in each lake) out of the whole training periods 30 times and tested the model performances in the fixed testing periods (Table 1). The results of these 30 times model runs were aggregated to assess the model performance in each lake.

To further evaluate model performance in detecting hypoxia, we define hypolimnetic hypoxia when bottom DO concentrations decreased below the specific thresholds. Given that the restoration actions have been taken in Furesø since 2003, hypolimnetic hypoxia has been reduced. Also, the sample interval of hypolimnetic DO in Lake Ekoln is over a month, and therefore not sufficient to interpolate the exact timing of anoxia. Thus, we only used Lake Erken, Müggelsee, and Lake Mendota, to evaluate model performance in detecting hypoxia events. DO <2 mg/L was used as the criterion for hypoxia in Lake Mendota which experiences serious eutrophication and anoxia in the seasonal stratified period (Nürnberg, 1995; Scavia et al., 2014), and the criterion was lifted to DO <3 mg/L in Lake Erken and Müggelsee since eutrophication and anoxia are not as serious as in Lake Mendota and the DO sampling intervals are higher in these two lakes (Howell & Simpson, 1994). We used the True Positive Rate (TPR; also refer as recall), and False Positive Rate (FPR) to identify the potential of ML models to accurately capture the hypoxia and risk of incorrectly send out the hypoxia warning (see Table S1 in Supporting Information S1). A model with 100% TPR and 0% FPR would constitute a perfect fit.

3. Results

3.1. Model Performance

Despite diversity in physical size, lake mixing regime and trophic state of the lakes tested in this study, all three model approaches simulated the seasonal variation of both surface and bottom DO well (see, Figures S2–S11 in Supporting Information S1). In Figures 2 and 3, R^2 and NMAE values of the 30 separate model runs were plotted

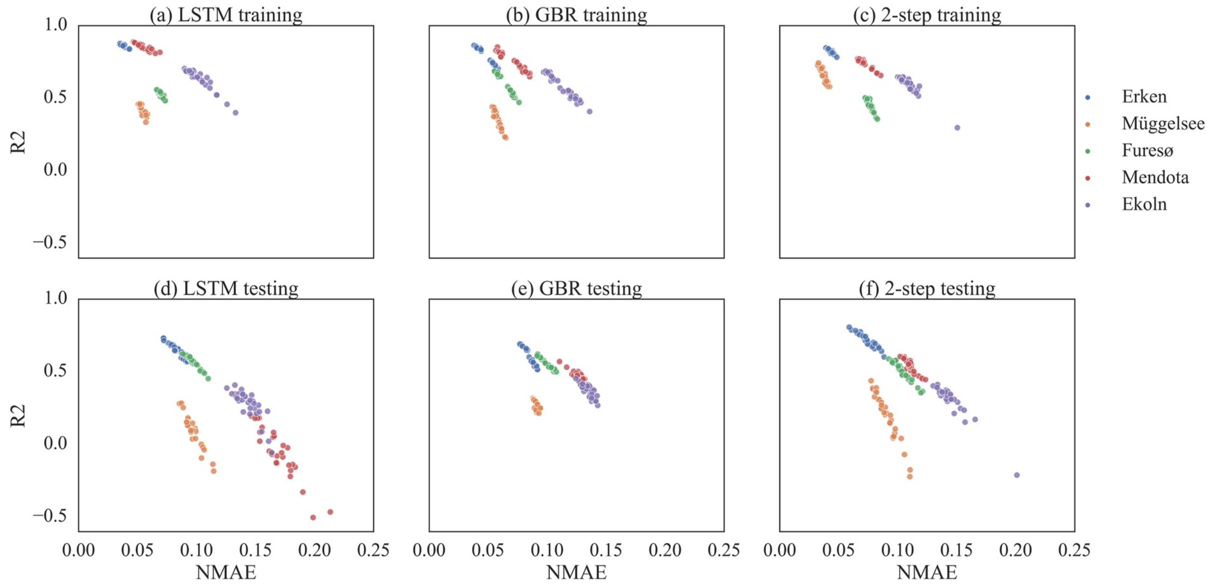

Figure 2. Evaluating metrics of three ML model approaches for DO in the surface water in panel (a–c) training data set, and (d–f) testing data set. Each point represents one model run, and the points located in the left upper corner of the figure means better model performance.

for each lake in order to illustrate the effects of variations in the training data set on the models' predictive power. For surface DO simulation in the testing periods, three approaches performed best in Lake Erken with direct LSTM and 2-step mixed model presenting averaged $R^2 > 0.6$ and averaged NMAE < 0.1, and GBR presenting averaged $R^2 > 0.5$ and averaged NMAE < 0.1. For bottom DO simulation in the testing periods, direct LSTM model shows best performance in Lake Erken with averaged $R^2 > 0.8$, and both direct GBR and 2-step mixed model presented best performance in Lake Mendota with averaged $R^2 > 0.8$ and NMAE < 0.1 mg/l.

The direct GBR approach showed more stable model performance with less variation in both the training and testing period (Figures 2 and 3, also see coefficient of variation in Supporting Information S1, Table S2) than the direct LSTM and 2-step mixed models. Combining GBR and LSTM into the 2-step mixed model improved the accuracy of both surface and bottom DO predictions in Mendota by increasing R^2 and decreasing NMAE.

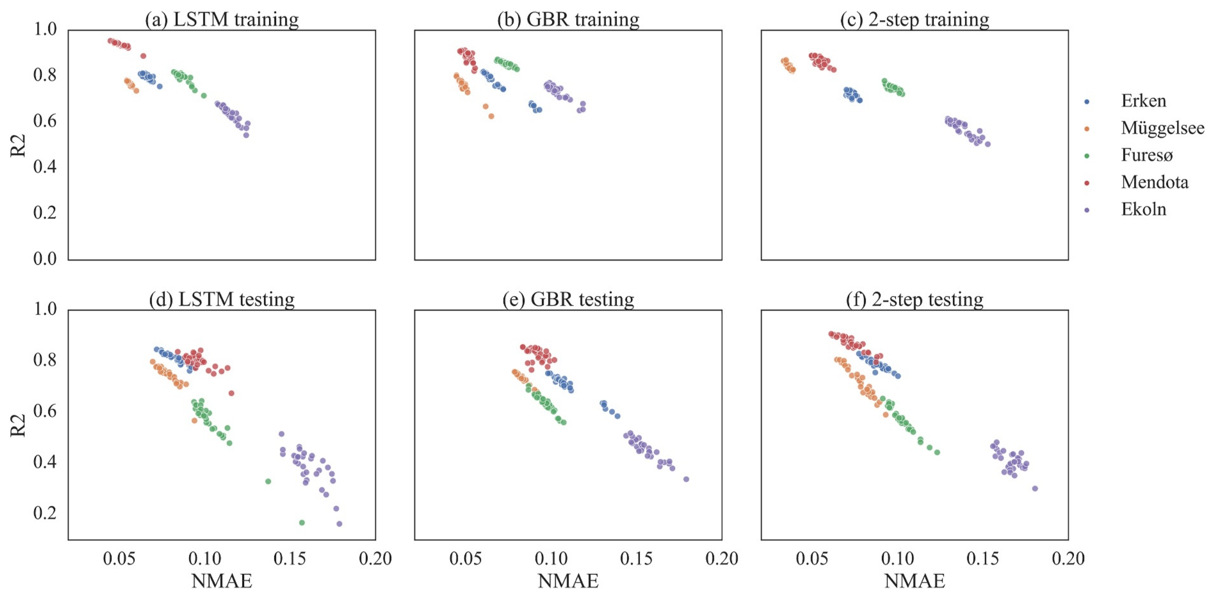

Figure 3. Similar as Figure 2, but for DO in the bottom water in (a–c) training data set, and (d–f) testing data set.

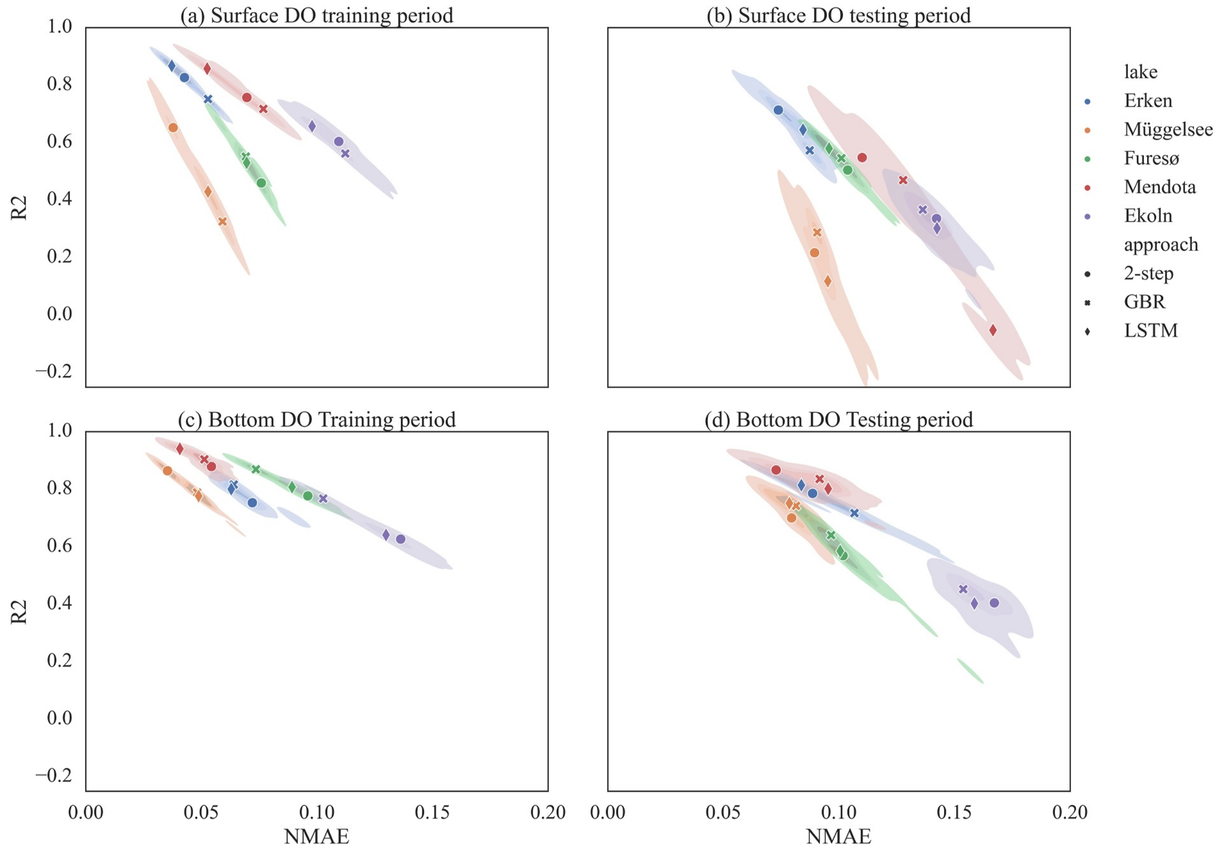

Figure 4. Performance of three ML model approaches. The shade areas represent the density plots of all the results from three approaches, and the symbols (i.e., ×, ◆, ●) represent the medians of R^2 and NMAE of three approaches.

Besides, it also decreased the variation of R^2 in the surface DO simulations in Lake Erken and the bottom DO simulation in Lake Mendota (Table S2 in Supporting Information S1). However, in some cases, the 2-step mixed model reduced model accuracy, for example, for surface DO simulation in Furesø.

Among the five lakes, the simulations of DO concentrations in Lake Ekoln were the worst with averaged $R^2 < 0.4$ for surface DO in the testing data and <0.6 in the training data, bottom DO had a R^2 ~0.5 in the testing data and <0.7 in the training data (Figure 4). Also, the modeled predictions had a larger variance than that in other lakes (Table S2 in Supporting Information S1). This presumably was due to the relatively large sampling interval (31 days) of DO even in this long data series, which suggests that the performance of ML models to some extent relies on the temporal resolution of training data.

In Furesø, most testing results showed that ML models learned the effect of oxidation treatment from training data and achieved generally promising R^2 and NMAE values. However, the oxidation treatment failed in 2015–2017 which was not captured by all three ML models (see, Figure S7 in Supporting Information S1). The over-simplification of the hypolimnetic oxidation to a single binary factor could be one of the reasons.

Overall, every tested ML model showed comparable evaluating metrics in both training and testing data sets (Figure 4), and the issue of overfitting did exist in Lake Ekoln, which has most sparse observations of DO, and surface DO prediction in Lake Mendota and Müggelsee. Direct LSTM models, compared to GBR, show more vulnerability to overfitting since tree models (e.g., GBR) can better address overfitting. To some extent, the 2-step mixed model further narrows down the difference in evaluating metrics between training and testing data compared to direct LSTM models (Figures 2 and 3 and Figures S2–S11 in Supporting Information S1).

Besides, in some model runs the testing data showed even higher MAE and RMSE (e.g., Surface DO in Lake Furesø, Müggelsee, Figure S6 in Supporting Information S1, Figure 8). This could be attributed to the relatively longer training period including the measurements from extreme observations (e.g., high values in the winter of

10.1029/2023EA003473

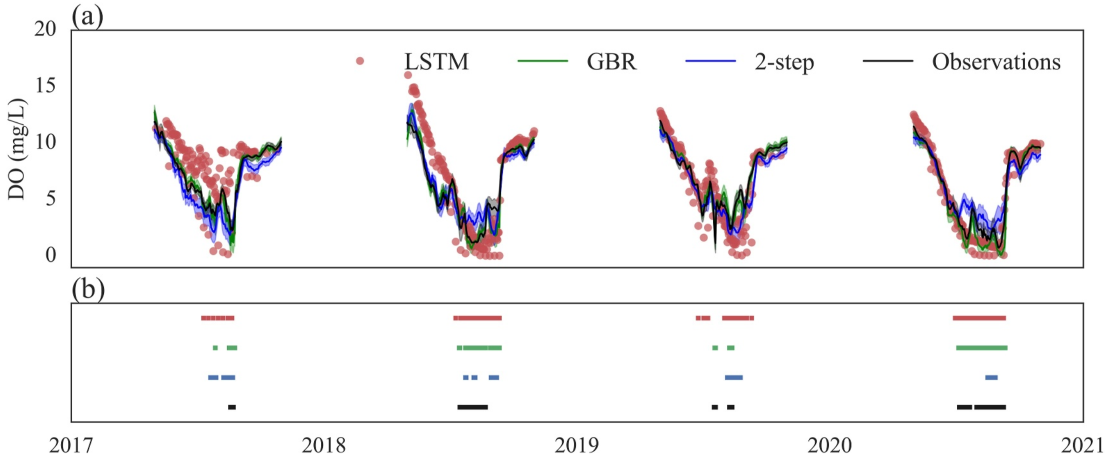

Figure 5. (a) Timeseries of YSI observed and modeled (averaged over 30-time model runs) bottom DO concentration in Lake Erken during May–September. The shaded areas represent the 95% confidence interval of each model approach. (b) Observed and modeled hypoxia events in the testing period.

1993 to 2000, and low values in the summer of 2005; Figure S6 in Supporting Information S1) which did not exist in the testing period. However, this unusual trend was eliminated when applied normalization to MAE (Figures 2 and 4).

3.2. Hypoxia Detection

All three model approaches captured the low bottom DO values during the stratified season (Figures 5–7). In Lake Erken and Müggelsee, the models were able to reproduce the trends in declining bottom DO concentration, but the magnitudes of simulated bottom DO decline were not enough to be counted as hypoxia events, and therefore counted as False Negative (FN; See Table S1 in Supporting Information S1) (e.g., the hypoxia in the July of 2019 in Lake Erken and in the June of 2020 in Müggelsee). In Müggelsee, the hypolimnetic hypoxia in the stratified season of 2020 was only captured by the 2-step mixed model, which also showed better representation of bottom DO variation in the whole testing period than direct LSTM and GBR approaches (Figure 6). The 2-step mixed model outperformed the other two approaches in Müggelsee with an average TPR over 70% and highest TPR closed to 90% (Figure 8b). In Lake Erken, direct LSTM outperformed the other two approaches, presenting averaged (~70%) and highest (~80%) TPRs of hypoxia detection (Figure 8a). In these two lakes, the FPRs of three approaches are below 10%, indicating that the possibilities of sending the wrong hypoxia warning are low (Figures 8d and 8e).

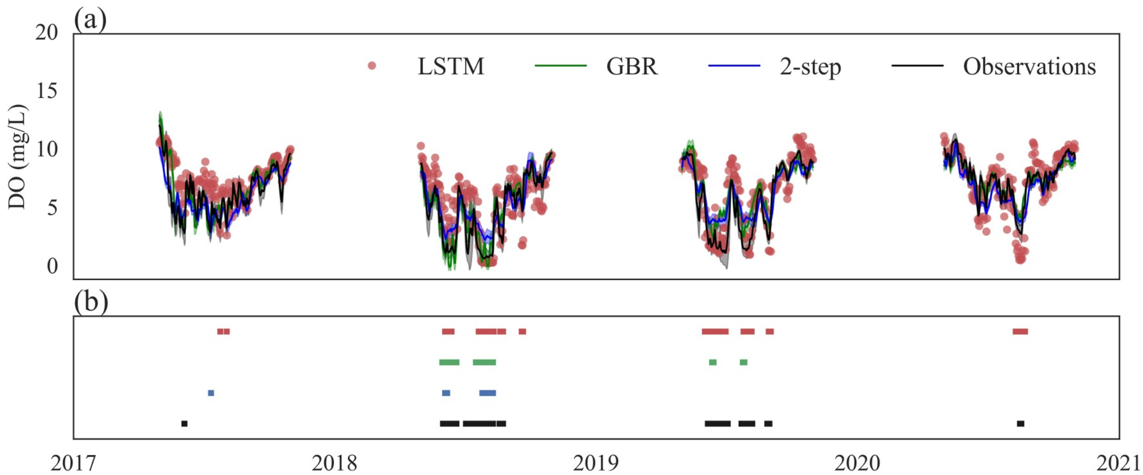

Figure 6. (a) Timeseries of YSI observed and modeled bottom DO concentration (May–September) in Müggelsee. The shaded areas represent the 95% confidence interval of each model approach. (b) Observed and modeled hypoxia events in the testing period.

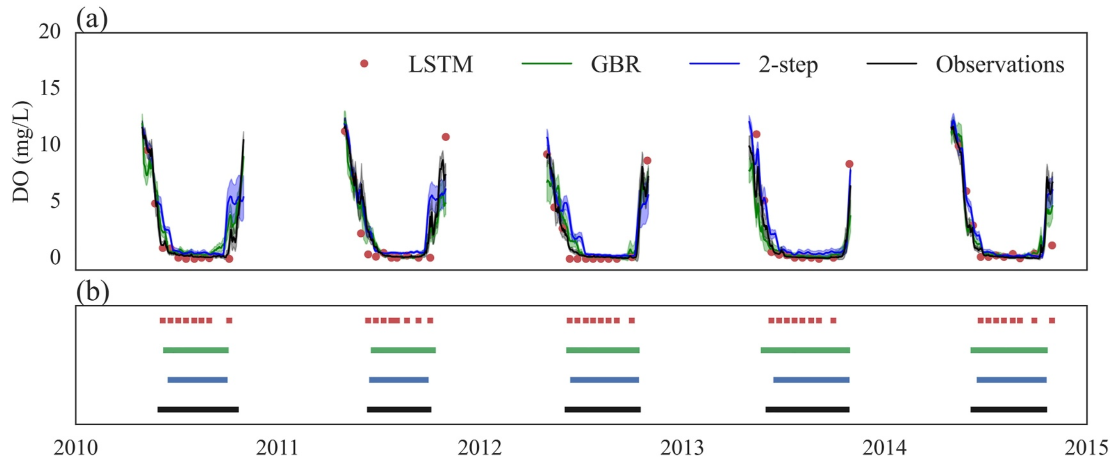

Figure 7. (a) Timeseries of YSI observed and modeled bottom DO concentration (May–September) in Lake Mendota. The shaded areas represent the 95% confidence interval of each model approach. (b) Observed and modeled hypoxia events in the testing period.

Compared to Lake Erken and Müggelsee, hypoxia at the bottom of Lake Mendota lasted almost the entire summer stratified period every year in the testing period. For this lake, the three ML approaches did not show any significant difference (Figure 7). The TPRs are relatively higher in Lake Mendota than the other two lakes with over 80% for all three ML approaches, presumably due to the longer and more stable hypoxia condition during stratified season in the bottom of Lake Mendota (Figures 7 and 8c). However, there is also a concomitant higher FPRs (over 30%) due to slight errors in the timing of the decline and rising in DO during spring and fall indicating the model is more likely to send an incorrect warning of hypolimnetic anoxia (Figure 8f) when exact timing is

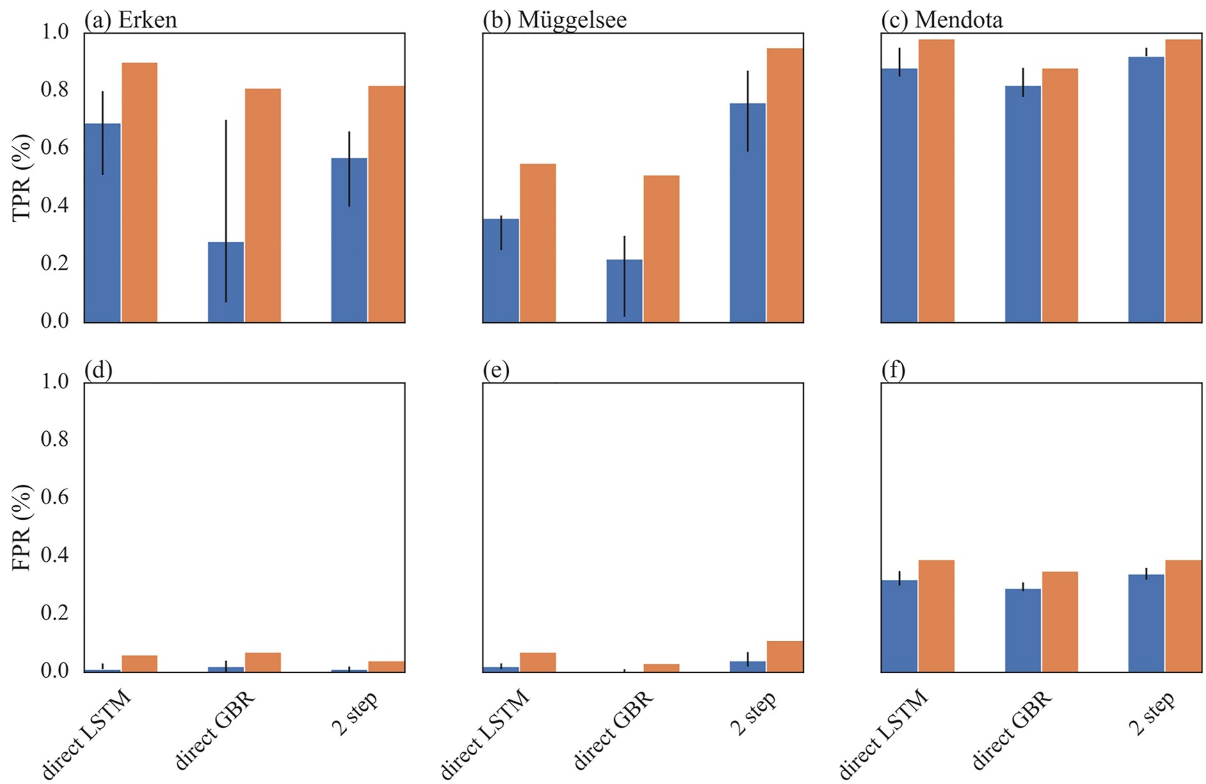

Figure 8. Evaluation of hypoxia detection in three lakes. The blue bars represent the TPRs and FPRs from the averaged hypolimnetic DO concentration predictions with error bars representing the 95% confidence interval, and the orange bars represent the TPRs and FPRs from minimal hypolimnetic DO concentration predictions over the 30-time model runs, indicating the highest hypoxia detection.

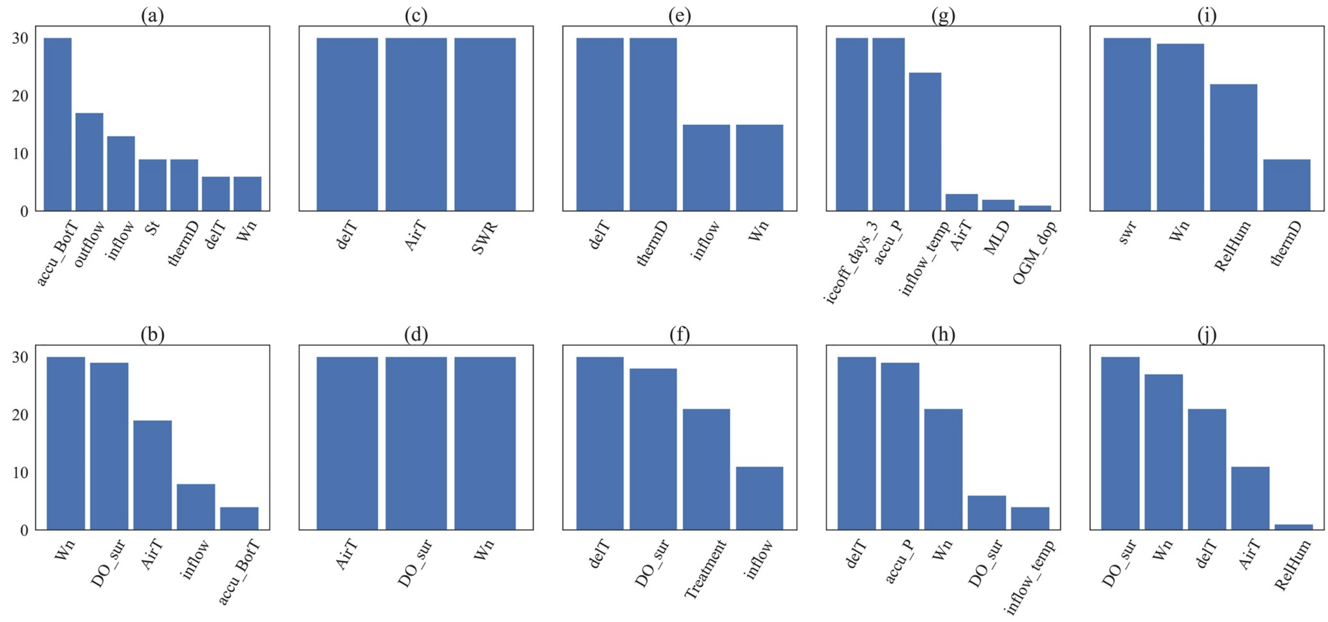

Figure 9. Top 3 important features in simulating surface (first row panels) and bottom DO (second row panels) in Lake (a, b) Erken, (c, d) Müggelsee, (e, f) Furesø, (g, h) Mendota, (i, j) Ekoln, *y*-axis shows the times of the feature was ranked as the top 3 features in the 30 test runs. The explanations of the short names can be found below, * accu_BotT: Accumulated bottom water temperature over 10 days, *St: Schmidt stability, *Wn: Wedderburn number, *delT: Temperature difference (delT) between surface water (averaged over the upper 3 m) and bottom water (bottom layers), *thermd: Thermocline depth, *AirT: Air temperature, *MLD: Mixing layer depth, *Ice-off date 3: Over 30 days from ice-off date, *accu_P: Accumulated phosphate from river loading, *inflow_temp: Water temperature of inflow, *OGM_dop: Concentration of dissolved organic phosphorus, *SWR: Shortwave radiation, *RelHum: Relative humidity. *Treatment: water treatment operation in Furesø.

critical. Overall, most of the hypoxic period is correctly predicted by all three ML approaches and the 2-step mixed model is more outstanding in predicting hypoxia than direct GBR and LSTM in the polymictic lake.

3.3. Feature Ranking

The GBR model can retrieve importance scores that indicates how useful each feature was in the construction of the boosted decision trees within the model, and it is computed explicitly for each individual decision tree by the amount that each attribute split point improves the performance measure, weighted by the number of observations the node is responsible for. At the end, the feature importances are averaged across all of the decision trees within the model. Thus, this additional benefit of using GBR could help to rank the dominant features controlling DO concentration variations and hypoxia events. Figure 9 summarizes the top three important features for surface and bottom DO concentration simulations from the 30 variable training data sets used for the GBR model runs. The results suggest that surface DO concentration is the essential factor for simulating bottom DO in most of the tested lakes, except Lake Mendota where surface DO only plays dominant roles in 5 out of 30 model runs, demonstrating the necessity of adding the predictive surface DO into training features of bottom DO. In Müggelsee, the only tested polymictic lake, the top 3 important features in simulating surface and bottom DO are consistent over the 30 test runs. The value of the hybrid modeling approach that makes use of information from the process-based models (Figure 1) is demonstrated by the importance of the derived hydrothermal variables for the prediction of DO. W_n, described as the ratio between the wind friction and the gradient of pressure established by the stratification (Patterson et al., 1984), is the dominant feature in bottom DO simulation except in Furesø which has an anthropogenic disturbance in the bottom water environment due to reaeration. Also, delT, which indicates the intensity of stratification played a major role in predicting the bottom DO in three out of five lakes. In Lake Mendota, delT and W_n play the major roles in controlling bottom DO, indicating that hypoxia was largely regulated by stratification dynamics (Ladwig et al., 2021a, 2021b). In addition, accumulated phosphate from river loading (accu_P) is also one of the top features for both surface and bottom DO in Lake Mendota, also demonstrating that ML models could account for relationships between DO concentration and external loading of nutrients without explicitly specifying the detailed biogeochemical relationships that would be needed in a process-based water quality model.

4. Discussion

4.1. Model Performance

The mechanisms that lead to variations in lake DO concentration and hypolimnetic hypoxia are complex and vary from lake to lake. This study tested the performance of ML approaches in simulating surface and bottom DO concentration in various lake systems via multiple environmental inputs. In addition to meteorological and hydrological inputs, hydrodynamic and ice-related variables, as well as other external disturbance factors (e.g., water treatment action, species invasion) have been considered in the model training. The accuracies of ML models surpassed results obtained from earlier studies using process-based models. Mesman et al. (2022) reported the results from General Ocean Turbulence Model (GOTM) in Lake Erken with RMSE for the full water column DO around 2.1 mg/L, while RMSEs of 2-step mixed ML approach were less than 1 mg/L in the surface water and less than 2 mg/L in the bottom water of Lake Erken. Ladwig et al., 2021a, 2021b applied GLM-AED2 model in Lake Mendota, showing RMSE in the surface layer is 2.77 mg/L and in the bottom layer is 3.31 mg/L, while the 2-step mixed ML approach has RMSE less than 2 mg/L in both surface and bottom layer. The polymictic characteristics of Müggelsee make the prediction of hypoxia more variable and therefore more challenging. Here, the 2-step mixed model workflow stands out among the three ML approaches, showing its advantages in this polymictic lake by successfully capturing the fluctuations in bottom DO concentrations that were recorded by the high-frequency YSI sensor.

The results from process-based modeling (Ladwig et al., 2021a, 2021b) revealed that external nutrient loading has a minor effect on the onset and duration of anoxia in the hypolimnion of eutrophic lake like Lake Mendota. However, the feature ranking from our study shows that accumulated external loading of Phosphate (i.e., accu_P) regulates both surface and bottom DO concentration here (Figure 9). These results do not completely contradict each other, since even though external nutrient loading may not affect the overall duration of anoxia, it may still play a role in controlling the variations in DO throughout the ice-free period.

Further, by accounting for external factors, like water treatment operation in Furesø as binary training features, three ML approaches can reproduce recovery of hypoxia in the bottom waters of a lake adapting to the treatment action (See Figure S7 in Supporting Information S1). When we excluded the water treatment feature from Furesø models (Figure S12 in Supporting Information S1), and all three approaches show obvious deficiency in capturing the bottom DO concentration. The training data in Furesø spans 20 years (1990–2009) with 7 years (2003–2009) occurring during oxidation treatment. Such a long historical data series of conditions with and without oxidation treatment provided the models with sufficient training so that the binary classification allowed the ML model, especially GBR, to learn the pattern (Breiman, 1984).

The observed response of DO to species invasion (and subsequent food web alterations) in Lake Mendota and Müggelsee are not as clear as the oxidation treatment effects in Furesø. The high frequency observations in Müggelsee did not show a clear trend of changing surface DO in response to the invasive species (Figure S8 in Supporting Information S1). We did observe a slight increase in the lowest bottom DO concentration in Müggelsee during 2011–2017 when Nuttall's waterweed dominated the macrophytes species. None of the three ML approaches were able to clearly capture this minor trend, but they did simulate the slight relief of hypoxia in 2014–2017 (Figure S9 in Supporting Information S1). The effect of invasive species involves much more complex ecological interactions which made it difficult for ML models to capture its overall impact on DO. Further, by expressing the invasive species response as binary variable masks the seasonal shifts in biomass occurring across different trophic states, that is, phytoplankton and zooplankton, in each lake's food web.

4.2. Hybrid Model

Successful ML predictions depend on the availability of long-term high-frequency data sets that can serve as training features for the ML algorithms. This can limit training data to measurements of meteorology and stream discharge which are routinely available at a daily measurement frequency and have been collected over long historical periods. Here we demonstrate a hybrid modeling approach that uses these same model inputs to first force a simple 1-D PB hydrodynamic model that in turn provides additional information describing the thermal structure and mixing dynamics of lakes. This hybrid model workflow therefore preprocesses available environmental inputs to provide additional information that is known to influence the ecology and biogeochemistry of lakes, and which has been demonstrated to improve algal bloom prediction in Lake Erken (Lin et al., 2023).

Unlike the knowledge-/physics-guided machine learning models proposed by Read et al. (2019), Jia et al. (2021) and Daw et al. (2022) which encoded the general physical relationship into ML model codes or used PB model results to pre-train ML models, our approaches directly use the physical variables generated by PB models as training features. Although inherent model bias could exist in these physical variables due to approximations and imperfect parameterizations, they still reveal the changing trend of environment and are critical predictors for DO.

In this study, hydrodynamic training features (e.g., W_n and delT) were found to be of importance for the prediction of DO in all lakes studied (Figure 9). This was particularly true for the bottom DO that would be of greatest interest for water management. Here, increased water column stability limits vertical fluxes of DO and nutrients in the lake, further limiting deep-water DO renewal. Eventually, DO sinks dominate the mass budget in hypolimnion resulting in hypoxic conditions.

4.3. 2-Step Mixed Model Workflow

In principle, carrying the memory of inputs of the previous week should allow the LSTM model to better represent hypolimnetic DO depletion (Foley et al., 2012). However, due to the more complicated model architecture, it takes more computational resources to train LSTM model than GBR model, and the model training time is highly dependent on the length of training data and the number of features. We found that the design of 2-step mixed model workflow that first prescreened the training features for the second step LSTM model using the feature ranking from the GBR model reduces overall computational costs. Even more importantly, the prescreening leads to an overall improvement in the LSTM performance as can be seen by the difference in the R^2 values between direct LSTM and the 2-step mixed model workflow results in Figures 2 and 3. The 2-step mixed model workflow which also uses the LSTM model showed better performance in detecting hypoxia events in all three tested lakes where high frequency measurements were available particularly for the polymictic Müggelsee (Figure 8). The accuracy of the 2-step approach is promising, with highest TPRs in Lake Erken, Mügelsee, and Mendota, of over 80%.

The feature ranking provided by GBR model in the mixed model workflow (Figure 9) can also support our conceptual understanding of the interactions between DO dynamics and physical or biogeochemical processes, to calibrate the process-based numerical models (Ladwig et al., 2021a, 2021b), and to better design process-based models for specific water systems (Cortés et al., 2021).

4.4. Model Limitations

The major features we used to train the ML models are physical factors (e.g., Wind speed, delT, thermD, etc.) which have previously been shown to largely explain the DO variations (Bouffard et al., 2013; Cortés et al., 2021), but very few of these are strongly related to external nutrients loading which could also have an important impact on oxygen depletion and water quality especially for shallow lakes (Wetzel et al., 2001). Predictions of surface DO could be affected by not accounting for oxygen depletion due to the oxidation of DOM or DO production due to photosynthesis. In Müggelsee and Lake Mendota, the two most eutrophic lakes, the accuracy of surface DO model was lower than bottom DO model. Presumably, this was related to the lack of training data that would be more directly related to the processes affecting metabolism, that is, phytoplankton community composition, turnover rates, and biomass changes.

Since the ML model approaches applied here take the lake as a horizontally uniform system, they only resolve the temporal variations of depth-discrete (i.e., surface and bottom) DO concentrations. However, this assumption may not hold in large water systems with complex transport processes, and the training features (e.g., meteorological inputs, MLD, thermD, etc.) at a single point may not be sufficient to simulate the hypoxia in the whole lake, especially when the hypoxia has the spatial variation and related to the horizontal water mass transportation (Valipour et al., 2021). One of our tested lakes, Lake Ekoln, is a basin of Lake Mälaren, the third largest lake in Sweden, and the application of the model in this lake is an example of applying the ML approach to large and complex water system. In this case, training variables representing the circulation pattern or water mass exchange between the target region and other parts of the lake may be required if the interactions among the regions influence hypoxia events in different parts of the lake. Not only the local meteorological conditions, but the meteorological conditions affecting other parts of the large lake could lead to bottom water mass exchange and further trigger hypoxia in our region of interest, and should therefore be considered (Jabbari et al., 2019; Rao et al., 2008).

4.5. Model Applications

The purpose of this study was to evaluate the possibilities of using ML models to predict concentrations of lake DO, especially hypolimnetic DO in different lakes using only readily available measurements of meteorology and hydrology (Table 1). We demonstrate that this is possible. The three ML approaches tested in the study are shown to be powerful tools for reproducing and predicting DO concentrations, opening the possibility for similar algorithms to be incorporated into forecasting workflows that would predict lake DO, and the onset loss and duration of hypoxia events. Such a forecasting system could be the cost-effective choice for early warning and short-term forecast of anoxia events, supporting the decision making in drinking water plants, or providing estimates of potential fish habitat loss and internal nutrient loading (Nürnberg et al., 2013; Orihel et al., 2017). In the long term, warmer lake surface temperatures and stronger stratification are obvious effects of climate warming for numerous lakes across the world, which further suggest potential increases of the occurrence, duration, and extent of hypoxia in the hypolimnion (Ladwig et al., 2021a, 2021b; North et al., 2014). Our results demonstrate that ML approaches could also play a role in projecting DO under future climate scenarios (Jane et al., 2021). Our results highlight that surface DO concentration is an important feature for predicting bottom DO and hypoxia in our workflow (Figure 9). Thus, predicting surface and bottom DO in sequence with the former variable serving as one of the model inputs of the latter one can potentially improve the accuracy of hypoxia detection and prediction in future forecast systems.

Data Availability Statement

Model version 1.0.0 has been archived under https://doi.org/10.5281/zenodo.7613549, and it is also available at https://github.com/Shuqi-Lin/Dissolved-Oxygen-MLPrediction.git (Lin et al., 2024). All data from this study have been archived with the code in the "Training data" folder in the format used in the model. Data of Lake Erken were collected by the Erken laboratory in the archived format used by the Swedish Infrastructure for Ecosystem Science (SITES), and are available from the SITES data archive at https://hdl.handle.net/11676.1/qZYc4CMTOyxgvjv_gTAW08SO, Erken Laboratory (2022) (Lake Erken, last access: September 2022). Data of Lake Mendota were archived in Environmental Data Initiative, https://doi.org/10.6073/pasta/418bf748dc2351f026c25111f7cbfd7e (Ladwig et al., 2021a, 2021b).

Acknowledgments

S.L., DP and this study are funded by the EU and FORMAS project 2018-02771, in the frame of the collaborative international Consortium BLOOWATER (https://www.bloowater.eu/) financed under the ERA-NET WaterWorks2017 Cofounded Call. This ERA-NET is an integral part of the 2018 Joint Activities developed by the Water Challenges for a Changing World Joint Program Initiative (Water JPI). BMK was funded by 2017–2018 Belmont Forum and BiodivERsA joint call for research proposals, under the BiodivScen ERA-Net COFUND program, and with the funding organizations German Science Foundation (AD 91 CE/22-1). The data of each lake were collected via GLEON Signal Processing Working Group (https://gleon.org/). We thank Dr. Benjamin Kraemer, Dr. Robert Ladwig, Dr. Fenjuan Rose Schmidt Hu for providing the data to this study.

References

Bennett, E. M., Reed-Andersen, T., Houser, J. N., Gabriel, J. R., & Carpenter, S. R. (1999). A phosphorus budget for the lake Mendota watershed. *Ecosystems*, *2*(1), 69–75. https://doi.org/10.1007/s100219900059

Bouffard, D., Ackerman, J. D., & Boegman, L. (2013). Factors affecting the development and dynamics of hypoxia in a large shallow stratified lake: Hourly to seasonal patterns. *Water Resources Research*, *49*(5), 2380–2394. https://doi.org/10.1002/wrcr.20241

Breiman, L. (1984). *Classification and regression trees* (1st ed.), Routledge. https://doi.org/10.1201/9781315139470

Burchard, H., Bolding, K., & Villarreal, M. R. (1999). *GOTM, a General Ocean Turbulence Model: Theory, implementation and test cases* (Vol. 103). Joint Research Centre, Space Applications Institute.

Carpenter, S. R., & Kitchell, J. F. (1988). Consumer control of Lake productivity. *BioScience*, *38*(11), 764–769. https://doi.org/10.2307/1310785

Charlton, M. N., & Lean, D. R. S. (1987). Sedimentation, resuspension, and oxygen depletion in Lake Erie (1979). *Journal of Great Lakes Research*, *13*(4), 709–723. https://doi.org/10.1016/S0380-1330(87)71685-2

Cortés, A., Forrest, A. L., Sadro, S., Stang, A. J., Swann, M., Framsted, N. T., et al. (2021). Prediction of hypoxia in eutrophic Polymictic lakes. *Water Resources Research*, *57*(6), e2020WR028693. https://doi.org/10.1029/2020WR028693

Daw, A., Karpatne, A., Watkins, W. D., Read, J. S., & Kumar, V. (2022). Physics-Guided Neural Networks (PGNN): An application in Lake temperature modeling. In *Knowledge-guided machine learning: Accelerating discovery using scientific knowledge and data* (1st ed.). Chapman and Hall/CRC.

Dehghani, R., Torabi Poudeh, H., & Izadi, Z. (2022). Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. *Model. Earth System. Environment*, *8*(2), 2599–2613. https://doi.org/10.1007/s40808-021-01253-x

Erken Laboratory. (2022). Meteorological data from Erken, Malma Island, 1988-10-12–2021-12-31 [Dataset]. *Swedish Infrastructure for Ecosystem Science (SITES)*. https://hdl.handle.net/11676.1/qZYc4CMTOyxgvjv_gTAW08SO

Farrell, K. J., Ward, N. K., Krinos, A. I., Hanson, P. C., Daneshmand, V., Figueiredo, R. J., & Carey, C. C. (2020). Ecosystem-scale nutrient cycling responses to increasing air temperatures vary with lake trophic state. *Ecological Modelling*, *430*, 109134. https://doi.org/10.1016/j.ecolmodel.2020.109134

Foley, B., Jones, I. D., Maberly, S. C., & Rippey, B. (2012). Long-term changes in oxygen depletion in a small temperate lake: Effects of climate change and eutrophication. *Freshwater Biology*, *57*(2), 278–289. https://doi.org/10.1111/j.1365-2427.2011.02662.x

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Goedkoop, W., Naddafi, R., & Grandin, U. (2011). Retention of N and P by zebra mussels (Dreissena Polymorpha Pallas) and its quantitative role in the nutrient budget of eutrophic Lake Ekoln, Sweden. *Biological Invasions*, *13*(5), 1077–1086. https://doi.org/10.1007/s10530-011-9950-9

Gurkan, Z., Zhang, J., & Jørgensen, S. E. (2006). Development of a structurally dynamic model for forecasting the effects of restoration of Lake Fure, Denmark. *Ecological Modelling*, *197*(1–2), 89–102. https://doi.org/10.1016/j.ecolmodel.2006.03.006

Hanson, P. C., Stillman, A. B., Jia, X., Karpatne, A., Dugan, H. A., Carey, C. C., et al. (2020). Predicting lake surface water phosphorus dynamics using process-guided machine learning. *Ecological Modelling*, *430*, 109136. https://doi.org/10.1016/j.ecolmodel.2020.109136

Heddam, S., & Kisi, O. (2018). Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, *559*, 499–509. https://doi.org/10.1016/j.jhydrol.2018.02.061

Hipsey, M. R., Bruce, L. C., Boon, C., Busch, B., Carey, C. C., Hamilton, D. P., et al. (2019). A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON), geosci. *Model Dev.*, *12*(1), 473–523. https://doi.org/10.5194/gmd-12-473-2019

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Howell, P., & Simpson, D. (1994). Abundance of marine resources in relation to dissolved oxygen in long Island sound. *Estuaries*, *17*(2), 394–402. https://doi.org/10.2307/1352672

Jabbari, A., Ackerman, J. D., Boegman, L., & Zhao, Y. (2019). Episodic hypoxia in the Western Basin of Lake Erie. *Limnology & Oceanography*, *64*(5), 2220–2236. https://doi.org/10.1002/lno.11180

Jane, S. F., Hansen, G. J. A., Kraemer, B. M., Leavitt, P. R., Mincer, J. L., North, R. L., et al. (2021). Widespread deoxygenation of temperate lakes. *Nature*, *594*(7861), 66–70. https://doi.org/10.1038/s41586-021-03550-y

Jia, X., Xie, Y., Li, S., Chen, S., Zwart, J., Sadler, J., et al. (2021). Physics-guided machine learning from simulation data: An application in modeling lake and river systems. In *Paper presented at 2021 IEEE International conference on Data Mining (ICDM)*.

Johansson, L. S., Søndergaard, M., & Andersen, P. M. (2021). Søer 2020. NOVANA. Aarhus Universitet, DCE – Nationalt center for Miljø og Energi, 80 s. - Videnskabelig rapport nr. *474*. http://dce2.au.dk/pub/SR474.pdf

Kakouei, K., Kraemer, B. M., & Adrian, R. (2022). Variation in the predictability of lake plankton metric types. *Limnology & Oceanography*, *67*(3), 608–620. https://doi.org/10.1002/lno.12021

Kakouei, K., Kraemer, B. M., Anneville, O., Carvalho, L., Feuchtmayr, H., Graham, J. L., et al. (2021). Phytoplankton and cyanobacteria abundances in mid-21st century lakes depend strongly on future land use and climate projections. *Global Change Biology*, *27*(24), 6409–6422. https://doi.org/10.1111/gcb.15866

Kisi, O., Alizamir, M., & Docheshmeh Gorgij, A. (2020). Dissolved oxygen prediction using a new ensemble method. *Environmental Science and Pollution Research*, *27*(9), 9589–9603. https://doi.org/10.1007/s11356-019-07574-w

Ladwig, R., Appling, A. P., Delany, A., Dugan, H. A., Gao, Q., Lottig, N., et al. (2022). Long-term change in metabolism phenology in north temperate lakes. *Limnology & Oceanography*, *67*(7), 1502–1521. https://doi.org/10.1002/lno.12098

Ladwig, R., Hanson, P. C., Dugan, H. A., Carey, C. C., Zhang, Y., Shu, L., et al. (2021a). Lake thermal structure drives inter-annual variability in summer anoxia dynamics in a eutrophic lake over 37 years ver 1 [Dataset]. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/418bf748dc2351f026c25111f7cbfd7e

Ladwig, R., Hanson, P. C., Dugan, H. A., Carey, C. C., Zhang, Y., Shu, L., et al. (2021b). Lake thermal structure drives inter-annual variability in summer anoxia dynamics in a eutrophic lake over 37 years. *Hydrology and Earth System Sciences*, *25*(2), 1009–1032. https://doi.org/10.5194/hess-25-1009-2021

Leon, L. F., Smith, R. E. H., Hipsey, M. R., Bocaniov, S. A., Higgins, S. N., Hecky, R. E., et al. (2011). Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. *Journal of Great Lakes Research*, *37*(1), 41–53. https://doi.org/10.1016/j.jglr.2010.12.007

Lin, S., Pierson, D., Ladwig, R., Kraemer, B., & Hu, F. R. S. (2024). Multi-model machine learning approach accurately predicts lake dissolved oxygen with multiple environmental inputs V1.0.0 [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.7613549

Lin, S., Pierson, D., & Mesman, J. (2023). Prediction of algal blooms via data-driven machine learning models: An evaluation using data from a well monitored mesotrophic lake, Geosci. *Model Dev. Discuss.*, *16*(1), 35–46. https://doi.org/10.5194/gmd-16-35-2023

Matsuzaki, S. I. S., Lathrop, R. C., Carpenter, S. R., Walsh, J. R., Vander Zanden, M. J., Gahler, M. R., & Stanley, E. H. (2021). Climate and food web effects on the spring clear-water phase in two north-temperate Eutrophic lakes. *Limnology & Oceanography*, *66*(1), 30–46. https://doi.org/10.1002/lno.11584

Mesman, J. P., Ayala, A. I., Goyette, S., Kasparian, J., Marcé, R., Markensten, H., et al. (2022). Drivers of phytoplankton responses to summer wind events in a stratified lake: A modeling study. *Limnology & Oceanography*, *67*(4), 856–873. https://doi.org/10.1002/lno.12040

Moras, S., Ayala, A. I., & Pierson, D. C. (2019). Historical modelling of changes in Lake Erken thermal conditions. *Hydrology and Earth System Sciences*, *23*(12), 5001–5016. https://doi.org/10.5194/hess-23-5001-2019

Müller, B., Bryant, L. D., Matzinger, A., & Wüest, A. (2012). Hypolimnetic oxygen depletion in Eutrophic lakes. *Environmental Science and Technology*, *46*(18), 9964–9971. https://doi.org/10.1021/es301422r

North, R. P., North, R. L., Livingstone, D. M., Köster, O., & Kipfer, R. (2014). Long-term changes in hypoxia and soluble reactive phosphorus in the hypolimnion of a large temperate lake: Consequences of a climate regime shift. *Global Change Biology*, *20*(3), 811–823. https://doi.org/10.1111/gcb.12371

Nürnberg, G. K. (1995). Quantifying anoxia in lakes. *Limnology & Oceanography*, *40*(6), 1100–1111. https://doi.org/10.4319/lo.1995.40.6.1100

Nürnberg, G. K., LaZerte, B. D., Loh, P. S., & Molot, L. A. (2013). Quantification of internal phosphorus load in large, partially polymictic and mesotrophic Lake Simcoe, Ontario. *Journal of Great Lakes Research*, *39*(2), 271–279. https://doi.org/10.1016/j.jglr.2013.03.017

Orihel, D. M., Baulch, H. M., Casson, N. J., North, R. L., Parsons, C. T., Seckar, D. C. M., & Venkiteswaran, J. J. (2017). Internal phosphorus loading in Canadian fresh waters: A critical review and data analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(12), 2005–2029. https://doi.org/10.1139/cjfas-2016-0500

Paerl, H. W., & Paul, V. J. (2012). Climate change: Links to global expansion of harmful cyanobacteria. *Water Research*, *46*(5), 1349–1363. https://doi.org/10.1016/j.watres.2011.08.002

Patterson, J. C., Hamblin, P. F., & Imberger, J. (1984). Classification and dynamic simulation of the vertical density structure of Lakes1. *Limnology & Oceanography*, *29*(4), 845–861. https://doi.org/10.4319/lo.1984.29.4.0845

Pierson, D. C., Pettersson, K., & Istvanovics, V. (1992). Temporal changes in biomass specific photosynthesis during the summer: Regulation by environmental factors and the importance of phytoplankton succession. *Hydrobiologia*, *243*(1), 119–135. https://doi.org/10.1007/BF00007027

Rao, Y. R., Hawley, N., Charlton, M. N., & Schertzer, W. M. (2008). Physical processes and hypoxia in the central Basin of Lake Erie. *Limnology & Oceanography*, *53*(5), 2007–2020. https://doi.org/10.4319/lo.2008.53.5.2007

Rao, Y. R., Howell, T., Watson, S. B., & Abernethy, S. (2014). On hypoxia and fish kills along the north shore of Lake Erie. *Journal of Great Lakes Research*, *40*(1), 187–191. https://doi.org/10.1016/j.jglr.2013.11.007

Read, J. S., Hamilton, D. P., Jones, I. D., Muraoka, K., Winslow, L. A., Kroiss, R., et al. (2011). Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environmental Modelling and Software*, *26*(11), 1325–1336. https://doi.org/10.1016/j.envsoft.2011.05.006

10.1029/2023EA003473

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. (2019). Process-guided deep learning predictions of Lake water temperature. *Water Resources Research*, *55*(11), 9173–9190. https://doi.org/10.1029/2019WR024922

Rohwer, R. R., Hale, R. J., Vander Zanden, M. J., Miller, T. R., & McMahon, K. D. (2023). Species invasions shift microbial phenology in a two-decade freshwater time series. *Proceedings of the National Academy of Sciences*, *120*(11), e2211796120. https://doi.org/10.1073/pnas.2211796120

Scavia, D., David Allan, J., Arend, K. K., Bartell, S., Beletsky, D., Bosch, N. S., et al. (2014). Assessing and addressing the re-eutrophication of Lake Erie: Central Basin Hypoxia. *Journal of Great Lakes Research*, *40*(2), 226–246. https://doi.org/10.1016/j.jglr.2014.02.004

Shatwell, T., & Köhler, J. (2019). Decreased nitrogen loading controls summer cyanobacterial blooms without promoting nitrogen-fixing taxa: Long-term response of a shallow lake. *Limnology & Oceanography*, *64*(S1), S166–S178. https://doi.org/10.1002/lno.11002

Thompson, R. O. R. Y., & Imberger, J. (1980). Response of a numerical model of a stratified lake to wind stress. In *Proc. Second Int. Symp. Stratified flows, IAHR, Trondheim, 1980* (pp. 562–570).

Valipour, R., León, L. F., Howell, T., Dove, A., & Rao, Y. R. (2021). Episodic nearshore-offshore exchanges of hypoxic waters along the north shore of Lake Erie. *Journal of Great Lakes Research*, *47*(2), 419–436. https://doi.org/10.1016/j.jglr.2021.01.014

Vilas, M. P., Marti, C. L., Adams, M. P., Oldham, C. E., & Hipsey, M. R. (2017). Invasive Macrophytes control the spatial and temporal patterns of temperature and dissolved oxygen in a shallow lake: A proposed feedback mechanism of macrophyte loss. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.02097

Walsh, J. R., Lathrop, R. C., & Vander Zanden, M. J. (2017). Invasive invertebrate predator, Bythotrephes Longimanus, reverses trophic cascade in a north-temperate lake. *Limnology & Oceanography*, *62*(6), 2498–2509. https://doi.org/10.1002/lno.10582

Wegner, B., Kronsbein, A. L., Gillefalk, M., van de Weyer, K., Köhler, J., Funke, E., et al. (2019). Mutual facilitation among invading Nuttall's waterweed and Quagga mussels. *Frontiers in Plant Science*, *10*, 789. https://doi.org/10.3389/fpls.2019.00789

Wetzel, R. G. (2001). *Limnology: Lake and river ecosystems*. Elsevier Science.

Wilson, H. L., Ayala, A. I., Jones, I. D., Rolston, A., Pierson, D., de Eyto, E., et al. (2020). Variability in epilimnion depth estimations in lakes. *Hydrology and Earth System Sciences*, *24*(11), 5559–5577. https://doi.org/10.5194/hess-24-5559-2020

Yousefi, A., & Toffolon, M. (2022). Critical factors for the use of machine learning to predict lake surface water temperature. *Journal of Hydrology*, *606*, 127418. https://doi.org/10.1016/j.jhydrol.2021.127418

Zhu, S., & Heddam, S. (2020). Prediction of dissolved oxygen in urban rivers at the three Gorges Reservoir, China: Extreme Learning Machines (ELM) versus Artificial Neural Network (ANN). *Water Qual. Res. J.*, *55*(1), 106–118. https://doi.org/10.2166/wqrj.2019.053

Ziyad Sami, B. F., Latif, S. D., Ahmed, A. N., Chow, M. F., Murti, M. A., Suhendi, A., et al. (2022). Machine learning algorithm as a sustainable tool for dissolved oxygen prediction: A case study of Feitsui Reservoir, Taiwan. *Scientific Reports*, *12*(1), 3649. https://doi.org/10.1038/s41598-022-06969-z

References From the Supporting Information

Environmental Data MVM. (2023). Swedish University of Agricultural Sciences (SLU). In *National data host lakes and watercourses, and national data host agricultural land*. Retrieved from https://miljodata.slu.se/mvm/

Magnuson, J. J., Kratz, T. K., & Benson, B. J. (2006). *Long-term dynamics of lakes in the landscape: Long-term ecological research on NORTH temperate lakes*. Long-Term Ecological Research.