International Journal of GEOMATE, June, 2026 Vol.30, Issue 142, pp.145-154
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2026.142.g15331
Geotechnique, Construction Materials and Environment

PHYSICS-INFORMED DEEP LEARNING FOR PREDICTING PHYTOPLANKTON DYNAMICS AND HYPOXIA IN ENCLOSED WATERS

Jinichi Koue[1], Sota Yoshikawa[1] and Katsutoshi Hirayama[1]

[1]Graduate School of Maritime Sciences, Kobe University, Japan

*Corresponding Author, Received: 01 July 2025, Revised: 02 April 2026, Accepted: 05 April 2026

ABSTRACT: Accurate forecasting of phytoplankton dynamics and hypoxia is essential for effective management of freshwater ecosystems under increasing climatic and anthropogenic pressures. This study develops a hybrid deep learning framework that integrates physico-chemical mass-balance equations with neural network models (NN, RNN, and LSTM) to predict chlorophyll-a, -b, -c, and dissolved oxygen (DO) in Lake Biwa, Japan. Three approaches are examined: Case 1 using conventional observational inputs, Case 2 incorporating balance-equation–derived mechanistic inputs, and Case 3 further introducing physics-informed constraints in the loss function for DO prediction. Short-term forecasts show that Case 2 improves the reproduction of chlorophyll peaks at 0.5 m depth, particularly for chlorophyll-b, with LSTM reducing RMSE from 0.41 to 0.35 μg/L. For DO, Case 2 consistently reduces prediction errors at 90 m depth across all models by approximately 15–25%. Case 3 yields additional short-term error reductions, although without statistical significance. Paired t-tests confirm significant short-term improvements for DO in the NN model ($p = 0.05$) and for chlorophyll-b in the NN model ($p = 0.02$). In long-term forecasting, chlorophyll prediction accuracy decreases in Case 2 due to fixed hyperparameter settings, whereas DO predictions show statistically significant improvements in both Case 2 and Case 3 ($p = 0.00$), highlighting the effectiveness of physics-informed constraints for capturing slow oxygen dynamics. Overall, the proposed framework provides a process-informed basis for adaptive water quality management and hypoxia mitigation in freshwater lakes.

Keywords: Deep Learning, Algal Blooms, Hypoxia, Lake Biwa, Ecosystem Modeling

1. INTRODUCTION

Climate change and anthropogenic pressures are increasingly disrupting the physical and biogeochemical stability of freshwater ecosystems worldwide. Among the most pressing consequences are the intensification of hypoxia and harmful algal blooms (HABs), which threaten aquatic biodiversity and compromise water resource sustainability. Despite long-term nutrient reduction efforts, large-scale cyanobacterial blooms persist in major lakes such as Lake Taihu in China [1] and Lake Erie in North America [2], underscoring the resilience of these phenomena to conventional mitigation. Similar signs of ecological degradation have also emerged in Lake Biwa, Japan's largest freshwater lake, driven by nutrient inflows, aquaculture effluents, and climate-induced changes in wind regimes and runoff patterns [3].

In recent years, coupled physical–biogeochemical models have advanced mechanistic understanding of lake ecosystems by explicitly representing nutrient cycling, phytoplankton dynamics, and oxygen consumption processes [3]. In parallel, data-driven machine learning approaches—particularly Long Short-Term Memory (LSTM) networks—have shown strong predictive capability in short-term forecasting of algal blooms and water quality variability [4]. However, most existing LSTM-based applications rely primarily on historical observations and meteorological forcings, treating lake ecosystems as black-box systems. Consequently, these models often suffer from limited physical and ecological interpretability, reduced extrapolation capability under non-stationary conditions, and insufficient robustness during extreme events such as bloom outbreaks or hypoxic collapse.

To overcome these limitations, hybrid modeling approaches that integrate process-based knowledge into machine learning frameworks have recently attracted increasing attention. In this study, the term *mechanistic* refers not to directly observed state variables alone, but to variables derived from governing equations that explicitly describe ecosystem processes. Specifically, outputs from mass-balance–based ecosystem equations—representing phytoplankton growth and loss processes, nutrient transformations, sediment oxygen demand, and biogeochemical feedbacks—are used as mechanistic inputs. These variables embed causal information on ecosystem functioning and are expected to enhance both the generalization performance and event-level reliability of data-driven models. Previous studies have demonstrated that purely data-driven models may achieve low average errors while failing to capture event peaks or abrupt

regime shifts, whereas hybrid approaches that incorporate mechanistic information tend to improve robustness during extreme events. Accordingly, recent hybrid modeling studies have emphasized the importance of event-oriented evaluation frameworks, such as threshold-based hit rates, event timing errors, and categorical performance measures, to better quantify practical forecasting utility. Physically interpretable empirical or hybrid models have been shown to improve predictions of phytoplankton dynamics in lake and reservoir systems influenced by water transfers [5]. Hybrid approaches integrating recurrent neural networks with physics-based models have also been reported to enhance the representation of temporal dynamics, supporting the adoption of event-based evaluation—such as peak magnitude and timing—alongside conventional error metrics in ecological forecasting [6].

Based on this concept, this study develops a hybrid forecasting framework that couples mechanistic ecosystem-model outputs with feedforward neural networks, recurrent neural networks (RNN), and Long Short-Term Memory (LSTM) models [7–10]. Long-term water quality observations from the Lake Biwa Environmental Research Center and meteorological forcings from the Japan Meteorological Agency are integrated with mechanistically derived variables to predict chlorophyll-a, chlorophyll-b, chlorophyll-c, and dissolved oxygen concentrations.

The objectives of this study are to: (i) evaluate the impact of incorporating mechanistic balance-equation outputs on the short- and long-term forecasting performance of deep learning models for water quality in Lake Biwa; and (ii) compare prediction accuracy between a purely data-driven approach (Case 1) and a hybrid approach that explicitly represents physico-chemical processes (Case 2), focusing on chlorophyll (a, b, c) and dissolved oxygen. In this study, although conventional statistical indices are used for quantitative comparison, particular attention is paid to model behavior during bloom peaks and hypoxia events. This perspective provides a complementary event-based interpretation of forecasting performance and supports the assessment of mechanistic-input hybrid models in an ecological management context.

2. RESEARCH SIGNIFICANCE

This study advances water quality forecasting by integrating physico-chemical process–based balance equations with deep learning models, thereby enhancing both predictive accuracy and practical decision-making value. Unlike conventional threshold-based alert systems, the proposed framework enables continuous, depth-resolved prediction of chlorophyll and dissolved oxygen, allowing earlier detection of bloom development and hypoxia risk. By providing process-informed forecasts rather than binary warnings, this approach supports proactive water treatment operations and adaptive ecosystem management, contributing to the advancement of AI-augmented ecological forecasting in freshwater environments.

3. METHODOLOGY

This study developed and compared two deep learning-based frameworks for forecasting water quality in Lake Biwa, utilizing both observational and process-informed input structures. Water quality data were obtained from the Lake Biwa Environmental Research Institute of Shiga Prefecture, and meteorological data were provided by the Japan Meteorological Agency. Two input configurations were considered: Case 1: Direct assimilation of observed physicochemical parameters. Case 2: Use of process-derived variables computed from a mass-balance ecosystem model originally developed by Koue [3], which represents key biogeochemical dynamics within the lake. These variables were designed to encode mechanistic insights into the data.

For both input cases, two types of deep learning architectures were employed: a Recurrent Neural Network (RNN), and a Long Short-Term Memory (LSTM) network. This comparative modeling approach enabled assessment of the predictive performance and interpretability associated with purely empirical versus process-based input features.

3.1 Machine Learning Predictive Technology

In this study, three neural network architectures-feedforward neural networks (NN), simple recurrent neural networks (RNN), and long short-term memory networks (LSTM)-are employed to forecast water quality variables in Lake Biwa, allowing systematic evaluation of temporal structure and mechanistic inputs under controlled model complexity. The feedforward NN serves as a baseline without explicit temporal dependence, whereas RNN and LSTM incorporate recurrent connections to represent time-series dynamics. For the simple RNN, the hidden state h_t at time t is updated as

$$h_t = act(Wx_t + Uh_{t-1} + b) \quad (1)$$

where x_t is the input vector, W and U are weight matrices, b is a bias term, and $act(\cdot)$ denotes a nonlinear activation function.

LSTM extends this structure by introducing memory cells and gating mechanisms to better capture long-term dependencies and mitigate vanishing gradient effects. It is used here to examine

whether enhanced memory representation improves forecasting performance relative to the simple RNN, particularly for long-term variability in chlorophyll and dissolved oxygen.

The simple RNN was intentionally selected to maintain architectural consistency among the examined models, limit parameter complexity given the available data length, and isolate the influence of mechanistic inputs rather than architectural sophistication. To ensure fair comparison, the same activation function and loss function were applied across all NN, RNN, and LSTM models. Specifically, the Rectified Linear Unit (ReLU) was used as the activation function in hidden layers, and the mean squared error (MSE) was adopted as the loss function during training. Differences in model performance are therefore attributed to architectural structure and the inclusion of mechanistic balance-equation outputs, rather than differences in optimization settings.

3.2 Deep Learning-Based Frameworks

The two frameworks developed in this study are used to predict water quality at depths of 0.5 m and 20 m, and 60 m at Imazu-oki in the northern part of Lake Biwa. The water quality data used as input data are twice-monthly observations by the Lake Biwa Environmental Research Institute of Shiga Prefecture, and the meteorological data are from the Hikone District Meteorological Observatory of the Japan Meteorological Agency.

The framework generates the concentration of chlorophyll a, b, and c (Chl-a, b, c), and dissolved oxygen (DO) as prediction output. Chlorophyll-a, -b, and -c are modeled separately rather than aggregated into total chlorophyll or functional groups. Chlorophyll-a represents total phytoplankton biomass, whereas chlorophyll-b and chlorophyll-c are associated with different phytoplankton communities, primarily green algae and diatoms (and other chromophytes), respectively. These pigments exhibit distinct seasonal variability, nutrient utilization, and vertical distribution patterns in Lake Biwa. Modeling each pigment individually allows the framework to better capture differential phytoplankton responses to environmental forcing and biogeochemical processes, thereby providing more detailed insight into bloom composition and ecosystem dynamics than total chlorophyll alone. In Case 1, observed environmental variables were used directly as model inputs. Table 1 shows the observation data used. Biweekly observational data were used, with missing values interpolated linearly and observation times standardized across variables prior to model input. In Case 2, model inputs were derived from a mass-balance formulation representing key biogeochemical processes. These values were calculated using a mechanistic water quality model previously developed by Koue [3], which simulates nutrient and phytoplankton dynamics in Lake Biwa. The outputs of the mass-balance equations are used directly as input variables for the deep learning models, without additional normalization or transformation beyond the preprocessing applied uniformly to all input data.

Concentration of phytoplankton (Chlorophyll-a, Chlorophyll-b, Chlorophyll-c) is calculated as follows:

$\frac{dP}{dt}$ = Photosynthetic growth + Phytoplankton mortality + Grazing by zooplankton + Respiration + Senescence + Sedimentation (2)

$\frac{\partial DO}{\partial t}$ = Photosynthetic production + Respiration by phytoplankton + Respiration by zooplankton + Oxygen consumption by particulate organic matter + Oxygen consumption by dissolved organic matter + Sediment oxygen demand + Reaeration + Oxygen consumption associated with nitrification (3)

Where, P represents the concentration of phytoplankton, and DO represents the concentration of dissolved oxygen.

In addition to Cases 1 and 2, a third framework (Case 3) is introduced for dissolved oxygen (DO) prediction only. Case 3 extends the hybrid input configuration of Case 2 by incorporating a physics-informed constraint into the loss function, resulting in a Physics-Informed Neural Network (PINN)–type formulation.

In Case 3, the same mechanistic input features derived from the mass-balance formulation used in Case 2 are employed. In addition, the loss function includes a constraint based on a simplified DO mass-balance model. Unlike the detailed balance equations used to compute the input features, the simplified model omits the oxygen consumption terms associated with zooplankton respiration and nitrification from Eq. (3). This simplification is adopted to improve numerical stability while retaining the dominant oxygen production and consumption processes governing bottom-layer DO dynamics.

To evaluate the mass-balance constraint at the prediction time, auxiliary variables required for the balance calculation-namely DO, water temperature, chlorophyll-a, dissolved total organic carbon (D-TOC), and particulate total organic carbon (P-TOC)-are included in the training dataset as values from the

previous time step (t−1).

The loss function for Case 3 is defined as

$$L = L_{data} + \lambda_{balance} \cdot L_{physics} \quad (4)$$

where L_{data} represents the data loss, defined as the mean squared error between the predicted and observed DO concentrations. The physics loss $L_{physics}$ quantifies the discrepancy between the DO prediction obtained from the simplified mass-balance model and that produced by the deep learning model. The weighting coefficient $\lambda_{balance}$, which controls the relative importance of the physical constraint, is optimized using Optuna within the range of 0.000 to 0.05. The parameters for each deep learning model are as Table 2.

Table 1 Observation data used as input values in Case 1

Predicted value	Input Parameters
Chl-a, b, c	Ammonium nitrogen, Nitrite nitrogen, Nitrate nitrogen, Phosphate phosphorus, pH, Water temperature, Transparency, DO
DO	Chlorophyll-a-c, Ammonium nitrogen, Nitrite nitrogen, Nitrate nitrogen, Phosphate phosphorus, pH, Water temperature, Transparency, DO, Dissolved organic

Table 2 The parameters for each deep learning model

(a) Short-term Forecasting Case

Case	Target	Model	Hidden Layers	Activation Function	Optimizer	Sequence Length	Batch Size	Epochs
Case 1	Chlorophyll a-c	NN	2 layers (64 units each)	ReLU	Adam	-	15	100
		RNN	1 layer (4 units)	tanh	Adam	24	15	100
		LSTM	1 layer (4 units)	tanh	Adam	24	15	100
	Dissolved Oxygen	NN	1 layers (22 units each)	ReLU	Adam	-	8	182
		RNN	1 layer (19 units)	tanh	Adam	24	8	127
		LSTM	1 layer (19 units)	tanh	Adam	24	8	127
Case 2	Chlorophyll a-c	NN	2 layers (64 units each)	ReLU	Adam	-	64	100
		RNN	1 layer (15 units)	tanh	Adam	24	64	100
		LSTM	1 layer (15 units)	tanh	Adam	24	64	100
	Dissolved Oxygen	NN	1 layers (11 units each)	ReLU	Adam	-	32	40
		RNN	1 layer (14 units)	tanh	Adam	24	16	74
		LSTM	1 layer (25 units)	tanh	Adam	24	32	85
Case 3	Dissolved Oxygen	NN	1 layers (10 units each)	ReLU	Adam	-	64	95
		RNN	1 layer (42 units)	tanh	Adam	24	8	14
		LSTM	1 layer (28 units)	tanh	Adam	24	8	35

(b) Long-term Forecasting Case

Case	Target	Model	Hidden Layers	Activation Function	Optimizer	Sequence Length	Batch Size	Epochs
Case 1	Chlorophyll a	RNN	1 layer (16 units)	tanh	Adam	24	32	100
		LSTM	1 layer (16 units)	tanh	Adam	24	32	100
	Chlorophyll b	RNN	1 layer (8 units)	tanh	Adam	24	15	100
		LSTM	1 layer (8 units)	tanh	Adam	24	15	100
	Chlorophyll c	RNN	1 layer (16 units)	tanh	Adam	24	15	100
		LSTM	1 layer (16 units)	tanh	Adam	24	15	100
	Dissolved Oxygen	RNN	1 layer (8 units)	tanh	Adam	24	32	357
		LSTM	1 layer (27units)	tanh	Adam	24	8	80
Case 2	Chlorophyll a	RNN	1 layer (20 units)	tanh	Adam	24	15	101
		LSTM	1 layer (25 units)	tanh	Adam	24	15	101
	Chlorophyll b	RNN	1 layer (16 units)	tanh	Adam	24	32	100
		LSTM	1 layer (16 units)	tanh	Adam	24	32	100
	Chlorophyll c	RNN	1 layer (16 units)	tanh	Adam	24	32	100
		LSTM	1 layer (16 units)	tanh	Adam	24	32	100
	Dissolved Oxygen	RNN	1 layer (15 units)	tanh	Adam	24	64	141
		LSTM	1 layer (8 units)	tanh	Adam	24	64	149
Case 3	Dissolved Oxygen	RNN	1 layer (28 units)	tanh	Adam	24	32	85
		LSTM	1 layer (26 units)	tanh	Adam	24	64	72

3.3 Learning and Forecasting Periods

To evaluate predictive performance across different ecological time scales, both short-term and long-term forecasting models were developed. The training periods were defined separately for each target variable based on data availability. For phytoplankton biomass, the model was trained using data from April 1998 to March 2019. For dissolved oxygen, the training period spanned from April 2004 to March 2019.

Two forecasting horizons were considered: Short-term models were designed to predict water quality variables up to two weeks in advance, with forecasts validated over a three-year period from April 2020 to March 2023. Long-term models were developed for one-year-ahead predictions, with validation performed using data from April 2019 to March 2020.

The same network architecture and hyperparameters were used for both the short-term and long-term forecasting models to enable a consistent comparison of model performance across different temporal scales.

3.4 Evaluation Metrics

3.4.1 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is a metric used to quantify the prediction error between observed and forecasted values. It is defined as:

$$\mathrm{RMSE} = \sqrt{\left(\frac{\Sigma(y_i - \hat{y}_i)^2}{\mathrm{n}}\right)} \quad (4)$$

where y_i represents the observed values, $\hat{y}_i$ represents the predicted values, and n is the number of data points.

3.4.2 Mean Bias (MB)

The Mean Bias (MB) is used to evaluate systematic error (bias) between predicted and observed values. It is defined as:

$$MB = \frac{1}{n}\sum(\hat{y}_i - y_i) \quad (5)$$

where y_i represents the observed values, $\hat{y}_i$ represents the predicted values, and n is the number of data points. A positive MB indicates consistent overestimation by the model, while a negative MB indicates underestimation.

3.4.3 Paired t-tests

The paired t-test is used to evaluate whether there is a significant difference between two sets of paired (dependent) observations. It is defined as:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \quad (6)$$

where $d_i = (x_{i1} - x_{i2})$ represents the difference between paired values, $\bar{d}$ is the mean of the differences, s_d is the standard deviation of the differences, and n is the number of pairs. A large absolute value of t indicates that the mean difference is unlikely to have occurred by chance. The corresponding p-value is then used to determine whether the difference is statistically significant.

4. RESULTS

4.1 Short-term Forecasting Results

4.1.1 Chlorophyll-a-c

Figure 1 presents the prediction results for chlorophyll-a, -b, and -c in Cases 1 and 2 at a depth of 0.5 m, with the corresponding evaluation metrics summarized in Table 3, where the p-values from the paired t-tests are also reported. The analysis indicates that Case 2 achieved better agreement between observed and predicted values at bloom peaks for chlorophyll-b, compared with Case 1. This improvement can be attributed to the incorporation of balance equations that explicitly represent the physicochemical processes governing chlorophyll dynamics, thereby allowing the model to learn from inputs with stronger mechanistic relevance. As a result, the model was able to more accurately capture rapid fluctuations in chlorophyll-b concentrations. Overall, the selective improvement in performance metrics validates the effectiveness of the proposed hybrid approach. In addition, the paired t-test results summarized in Table 3 show that the reduction in errors reached statistical significance for chlorophyll-b with the NN model ($p = 0.02$), while other cases exhibited p-values above the 0.05 threshold. Although these latter improvements were not statistically significant, the reductions in RMSE and MB demonstrate practically meaningful gains, suggesting that Case 2 enhances predictive capability for specific phytoplankton pigments even where strict statistical significance is not achieved. The improvement in Case 2 was more evident for chlorophyll-b than for chlorophyll-a. This likely reflects the stronger coupling between chlorophyll-b–associated phytoplankton dynamics and the mechanistic processes represented in the balance equations, whereas chlorophyll-a integrates multiple phytoplankton groups with heterogeneous responses.

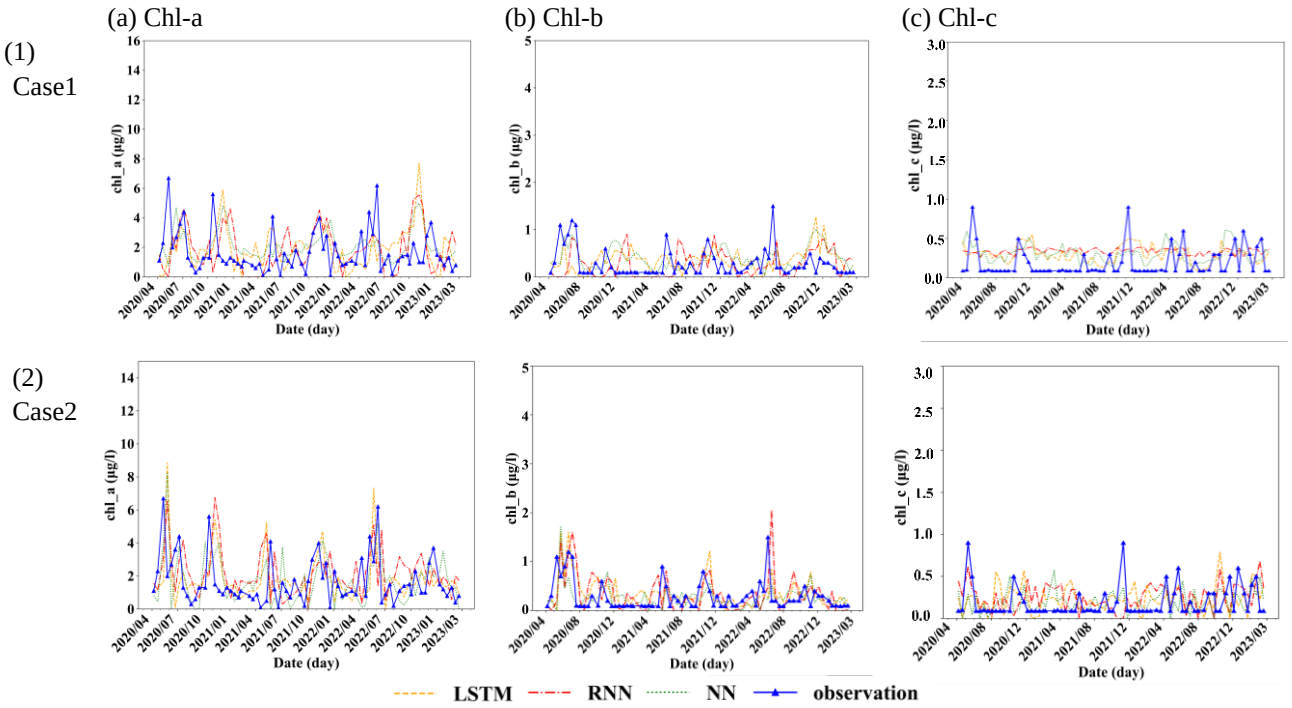


Fig. 1 Comparison of chlorophyll-(a) a, (b) b, (c) c, in (1) Case 1 and (2) Case 2 in the surface layer (0.5 m depth)

Table 3 Chlorophyll-a, b, and c evaluation index in the surface layer (0.5 m depth)

Predicted value	Model	RMSE (μg/l)		MB (μg/l)		p value
		Case1	Case2	Case1	Case2	
Chl-a	NN	1.67	1.67	0.44	0.06	0.25
	RNN	1.91	2.00	0.25	0.49	0.74
	LSTM	1.93	1.90	0.30	0.16	0.13
Chl-b	NN	0.35	0.35	-0.11	-0.02	0.02
	RNN	0.38	0.42	0.12	0.05	0.15
	LSTM	0.41	0.35	-0.06	0.06	0.07
Chl-c	NN	0.23	0.22	0.11	0.00	0.18
	RNN	0.24	0.24	0.15	0.06	0.38
	LSTM	0.23	0.24	0.11	0.04	0.33

4.1.2 Dissolved oxygen

To assess predictive performance under different input conditions, dissolved oxygen (DO) forecasts at a depth of 90 m were compared between Case 1 (conventional inputs), Case 2 (inputs derived from a balance equation), and Case3 (PINN-based framework). This depth, situated within the stratified hypolimnion of Lake Biwa's North Basin, is particularly vulnerable to oxygen depletion and nutrient accumulation, making it a critical zone for evaluating internal water quality degradation. Figure 2 illustrates the predicted DO dynamics for all cases, and the corresponding evaluation metrics are summarized in Table 4.

Compared with Case 1, Case 2 consistently reduced prediction errors across all models, with RMSE decreasing from 1.81 to 1.40 mg/L for the NN model, from 1.53 to 1.36 mg/L for the RNN model, and from 1.59 to 1.28 mg/L for the LSTM model. Case 3 further reduced RMSE values to 1.33 mg/L (NN), 1.28 mg/L (RNN), and 1.13 mg/L (LSTM), indicating an additional improvement in short-term predictive accuracy. However, statistical analysis reveals limitations in these improvements. Paired t-test results indicate that the differences in prediction accuracy are not statistically significant at the 5% level for any comparison (for Case 1–2 in the NN model, and for others).

Mean bias (MB) analysis highlights a shift in systematic error characteristics. While Case 1 and Case 2 generally exhibited positive biases (overestimation) in the RNN and LSTM models, Case 3 shifted toward negative biases (underestimation) across all models (e.g., -0.58 mg/L for NN and -0.49 mg/L for LSTM). This indicates that the strict physical constraints imposed by the PINN framework tend to suppress peak values, leading to a conservative prediction. Despite this shift in bias and the lack of statistical significance, the consistent reduction in RMSE underscores the practical utility of the PINN framework in minimizing total prediction error. The implications of this trend are further examined in the subsequent long-term forecasting section.

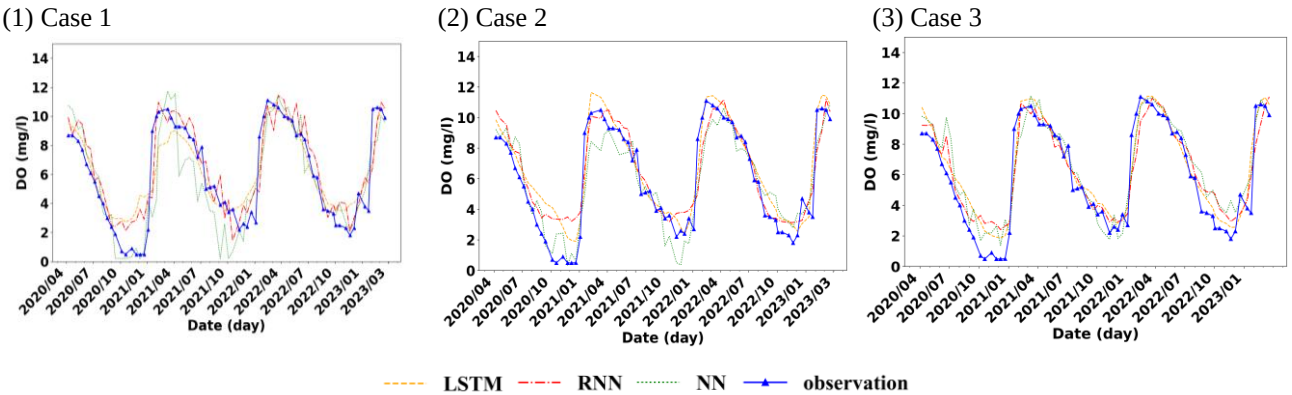


Fig. 2 Comparison of DO, in (1) Case 1, (2) Case 2, and (3) Case 3 in the bottom of the lake (90 m depth)

Table 4 DO evaluation index in the bottom of the lake (90 m depth)

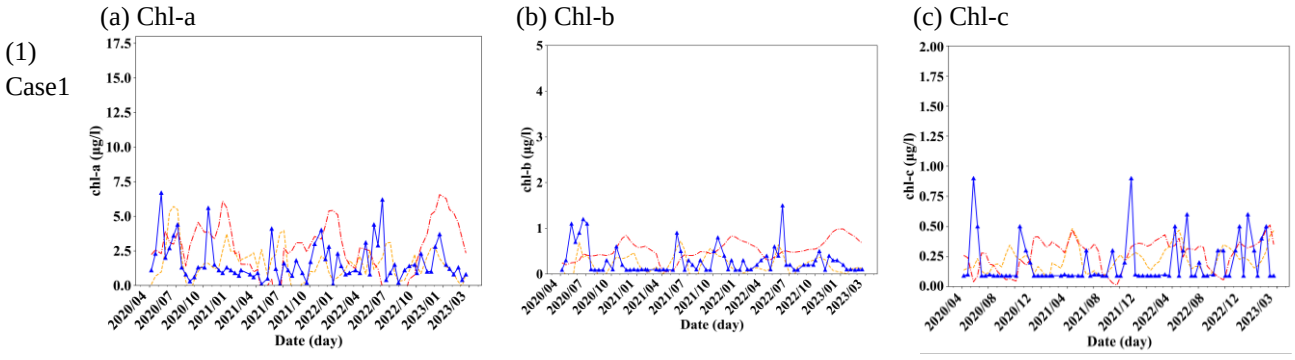
Predicted value	Model	RMSE (mg/l)			MB (mg/l)			p value	
		Case1	Case2	Case3	Case1	Case2	Case3	Case1-2	Case2-3
DO	NN	1.81	1.40	1.33	-0.43	-0.16	-0.58	0.07	0.61
	RNN	1.53	1.36	1.28	0.54	0.50	-0.38	0.12	0.34
	LSTM	1.59	1.28	1.13	0.25	0.65	-0.49	0.09	0.11

4.2 Long-term Forecasting Results

4.2.1 Chlorophyll-a-c

Figure 3 presents the predicted chlorophyll concentrations—(a) chlorophyll-a, (b) chlorophyll-b, and (c) chlorophyll-c—for Cases 1 and 2 in the surface layer (0.5 m depth), with the corresponding evaluation metrics summarized in Table 5. While Case 2 improved long-term prediction accuracy for chlorophyll-a, it exhibited reduced predictive accuracy for chlorophyll-b, with the decline most pronounced during bloom peaks.

Although Case 2 improved short-term forecasts, its inconsistent long-term performance across different chlorophyll components suggests that the chosen hyperparameters were insufficient to capture long-term temporal variability. In particular, for the RNN and LSTM models, hyperparameters such as the number of network layers, the number of neurons per layer, and the batch size failed to adequately represent the complex nonlinear processes underlying seasonal variability, species interactions, nutrient dynamics, and meteorological forcing. This insufficient parameterization likely contributed to the reduced forecasting accuracy observed in chlorophyll-b under Case 2. The paired t-test results in Table 5 show that Case 2 predictions significantly improved for chlorophyll-a (RNN: $p = 0.00$, LSTM: $p = 0.01$), whereas significant degradation was observed for chlorophyll-b (RNN/LSTM: $p = 0.00$), with MB shifting toward overestimation. For chlorophyll-c, RNN showed no significant difference ($p = 0.64$), while LSTM indicated a modest improvement ($p = 0.03$). Overall, these results suggest that although balance-equation inputs can enhance long-term prediction for some phytoplankton components (e.g., chlorophyll-a), inadequate hyperparameter settings limited their effectiveness for others, particularly chlorophyll-b.



(2) Case2

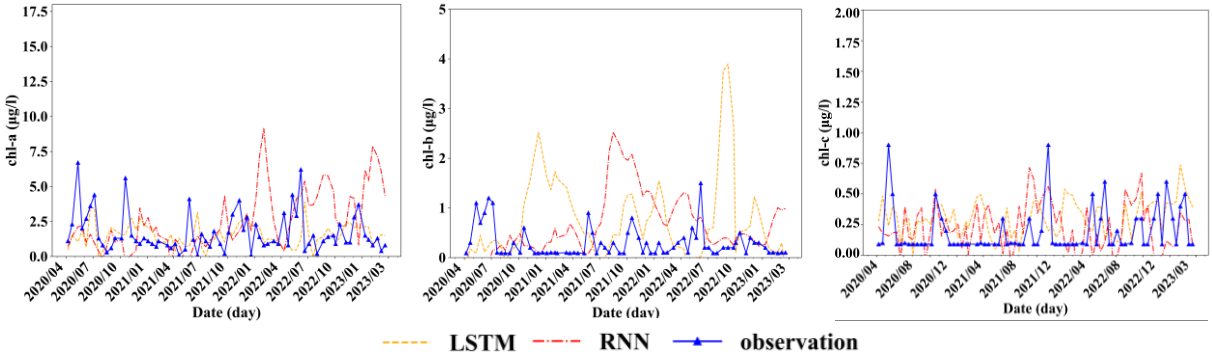


Fig. 3 Comparison of chlorophyll-(a) a, (b) b, (c) c, in (1) Case 1 and (2) Case 2 in the surface layer (0.5 m depth)

Table 5 Chlorophyll-a, b, and c evaluation index in the surface layer (0.5 m depth)

Predicted value	Model	RMSE (μg/l)		MB (μg/l)		p value
		Case1	Case2	Case1	Case2	
Chl-a	RNN	2.47	0.26	0.93	0.03	0.00
	LSTM	1.69	0.25	-0.06	0.12	0.01
Chl-b	RNN	0.57	0.82	-0.40	0.35	0.00
	LSTM	0.47	1.05	0.01	0.46	0.00
Chl-c	RNN	0.24	0.26	0.08	0.03	0.64
	LSTM	0.20	0.25	0.03	0.12	0.03

4.2.2 Dissolved oxygen

Figure 4 shows the predicted results of DO for Case 1, Case 2, and Case 3 at the lake bottom (90 m depth). Table 6 summarizes the evaluation indices for each model. The analysis indicates that while Case 1 failed to adequately reproduce the temporal variation patterns of dissolved oxygen concentration, Case 2 and Case 3 succeeded in capturing the observed dynamics with significantly higher accuracy.

Quantitative evaluation (Table 6) confirms a marked improvement in Case 2 compared to Case 1. For the RNN model, RMSE substantially decreased from 3.78 mg/L in Case 1 to 1.95 mg/L in Case 2. Similarly, for the LSTM model, RMSE was reduced from 2.33 mg/L to 1.91 mg/L. Paired t-test results confirmed that these error reductions from Case 1 to Case 2 were statistically significant for both models ($p = 0.00$).

Case 3 further improved predictive accuracy, achieving the lowest RMSE values of 1.83 mg/L (RNN) and 1.80 mg/L (LSTM). However, unlike the transition from Case 1 to Case 2, the improvement from Case 2 to Case 3 was not statistically significant ($p = 0.20$ for RNN, $p = 0.87$ for LSTM). This suggests that while the physics-informed framework consistently yields the best deterministic accuracy, the marginal gain over the balance-equation model (Case 2) varies statistically.

Mean bias (MB) analysis reveals a trade-off characteristic in the hybrid framework. Case 2 effectively reduced the positive bias (overestimation) observed in Case 1 across both models (e.g., from 0.99 to 0.54 mg/L in LSTM). In contrast, Case 3 exhibited an increase in positive bias (e.g., 1.14 mg/L in RNN), suggesting that the strict physics-based constraints might induce a systematic shift despite reducing the overall error variance. These findings demonstrate that balance-equation–derived inputs (Case 2) provide the primary contribution to long-term prediction accuracy under stratified conditions, while the additional constraints in Case 3 offer further refinement in RMSE, albeit with a trade-off in bias control.

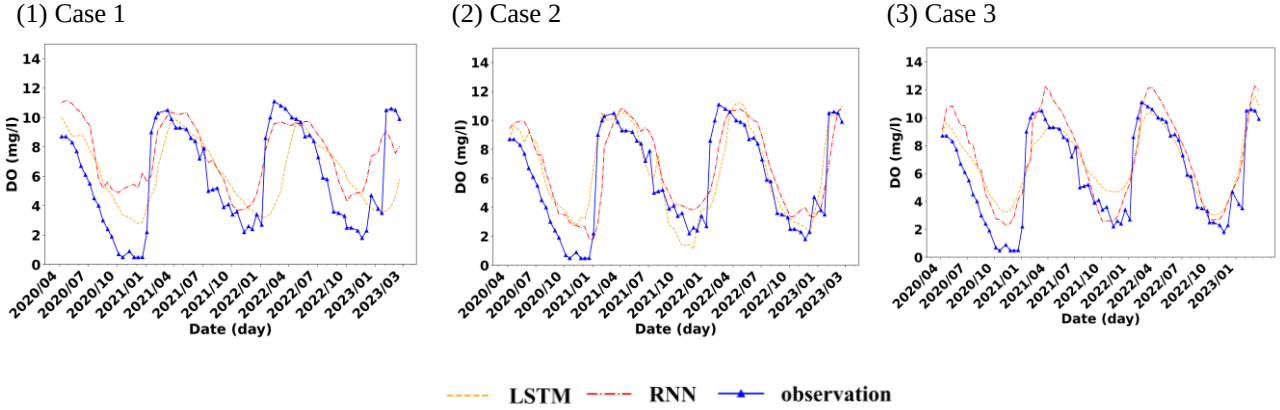


Fig. 4 Comparison of DO, in (1) Case 1, (2) Case 2, and (3) Case 3 in the bottom of the lake (90 m depth)

Table 6 DO evaluation index in the bottom of the lake (90 m depth)

Predicted value	Model	RMSE (mg/l)			MB (mg/l)			p value	
		Case1	Case2	Case3	Case1	Case2	Case3	Case1-2	Case2-3
DO	RNN	3.78	1.95	1.83	0.88	0.78	1.14	0.00	0.20
	LSTM	2.33	1.91	1.80	0.99	0.54	0.88	0.00	0.87

5. CONCLUSION

In this study, we compared the short-term and long-term predictions of water quality parameters in Lake Biwa among Case 1, which uses only conventional input values, Case 2, which incorporates balance-equation–derived inputs representing physico-chemical processes, and Case 3, which further introduces physics-informed constraints through the loss function for dissolved oxygen prediction. The results showed that Case 2 improved the prediction accuracy of chlorophyll (a, b, c) at the peak in the short-term prediction, confirming that the expression of physicochemical processes by the balance equation is effective. On the other hand, for dissolved oxygen, improvement was confirmed in Case 2 for all evaluation indices. This may be due to the fact that the introduction of the balance equation explicitly incorporated complex interactions such as oxygen consumption and production processes in the ecosystem into the model.

In the long-term prediction, a decrease in prediction accuracy was confirmed in Case 2 compared to Case 1 for all chlorophyll (a, b, c) species. This result may be due to the fact that the hyperparameters of the deep learning models, such as RNN and LSTM, do not properly capture long-term time-series variations. It should be noted that this reduced long-term prediction accuracy for chlorophyll reflects a limitation of the present framework, as hyperparameters were fixed across forecast horizons. Future work will address this issue by introducing horizon-dependent hyperparameter optimization to improve long-term forecasting performance, particularly for phytoplankton variables. For dissolved oxygen, Case 2 was able to capture the variability of observed values with high accuracy at the bottom of the lake (90 m depth). The paired t-test results further support these findings. Statistically significant improvements were confirmed for dissolved oxygen with the RNN and LSTM models ($p = 0.00$), and for chlorophyll-b with the NN model ($p = 0.02$). In contrast, the declines in chlorophyll-a and -b predictions under Case 2 were also statistically significant ($p < 0.05$), highlighting the limitations of hyperparameter settings in long-term forecasts.

Regarding Case 3, while the additional physics-informed constraint did not yield statistically significant improvements over Case 2 in terms of p-values ($p > 0.05$), it consistently achieved the lowest RMSE values for long-term dissolved oxygen prediction in both the RNN and LSTM models. This indicates that although the statistical difference is not significant, the physics-informed loss constraints are practically effective for minimizing error variance and capturing slow, cumulative oxygen dynamics under stratified conditions. These results confirm that the introduction of a balance equation has a markedly improved effect, especially in the short-term prediction of chlorophyll (a, b) and dissolved oxygen. The model developed in this study enables more accurate prediction of the timing of blue-green algae blooms and is expected to contribute to optimizing the timing of activated carbon application at water treatment plants. Furthermore, the demonstrated effectiveness of physics-informed constraints in long-term dissolved oxygen forecasting provides a foundation for future high-accuracy prediction of anoxia in Lake Biwa and for quantitatively assessing the contribution of dominant controlling processes.

6. ACKNOWLEDGMENTS

This work was partially supported by the Asahi Glass Foundation.

7. REFERENCES

[1] Cao H., Han L., Li L. A deep learning method for cyanobacterial harmful algae blooms prediction in Taihu Lake, China. Harmful Algae, Vol. 113, 2022, Article No.102189.

[2] Michalak A. M., Anderson E. J., Beletsky D. et al. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proceedings of the National Academy of Sciences, Vol. 110, Issue 16, 2013, pp.6448-6452.

[3] Koue J. Development of an ecosystem model considering sediment redox processes in enclosed water bodies. Water, Vol. 16, Issue 13,

2024, Article No.1879.

[4] Liang Z., Zou R., Chen X., Ren T., Su H., Liu Y. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. Journal of Hydrology, Vol. 581, 2020, Article No.124432.

[5] Fornarelli R., Galelli S., Castelletti A., Antenucci J. P., Marti C. L. An empirical modeling approach to predict and understand phytoplankton dynamics in a reservoir affected by interbasin water transfers. Water Resources Research, Vol. 49, Issue. 6, 2013, pp.3626-3641.

[6] Jia X., Willard J., Karpatne A., Read J., Zwart J., Steinbach M., Kumar V. Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. In Proceedings of the 2019 SIAM international conference on data mining, Society for Industrial and Applied Mathematics, 2019, pp.558-566.

[7] Elman J. L. Finding structure in time. Cognitive Science, Vol. 14, Issue 2, 1990, pp.179-211.

[8] Hochreiter S., Schmidhuber J. Long short-term memory. Neural computation, Vol. 9, Issue 8, 1997, pp.1735-1780.

[9] Gers F. A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM. Neural computation, Vol. 12, Issue 10, 2000, pp.2451-2471.

[10] Zhang S., Qi J. A study on the reliability of an artificial intelligence model for water environment evaluation. International Journal of GEOMATE, Vol. 25, Issue 107, 2023, pp.220-227.