Acta Oceanol. Sin., 2025, Vol. 44, No. 7, P. 147–160
https://doi.org/10.1007/s13131-025-2492-2

Enhanced forecasting of coastal chlorophyll-*a* through AdaBoost-optimized LSTM models

Wenxiang Ding[1, 2], Caiyun Zhang[1, 3*], Xueding Li[4], Liyu Zhang[5], Nengwang Chen[3]

[1] State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China
[2] Marine Science and Technology College, Zhejiang Ocean University, Zhoushan 316022, China
[3] Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, Xiamen University, Xiamen 361102, China
[4] Fujian Marine Forecasts, Fuzhou 350001, China
[5] Marine and Fishery Institute of Xiamen, Xiamen 361008, China

Received 29 October 2024; accepted 21 March 2025

Abstract

Algal blooms pose significant threats to marine ecosystems and human health. Accurate forecasting of chlorophyll-*a* (Chl-*a*) concentration is critical for effective control of harmful algal blooms (HABs). This study proposes a novel approach for enhancing Chl-*a* concentration forecasting by integrating the AdaBoost algorithm with long short-term memory (LSTM) neural networks. We developed a strong forecasting model by combining adaptive boosting (AdaBoost) with LSTM models in Xiamen Bay, China. This model achieved higher correlation coefficients and lower root mean square errors than individual weak models. The AdaBoost-optimized model increased the frequency of low absolute errors while decreasing the occurrence of high absolute errors, which indicated improved overall prediction accuracy and reliability. Moreover, the model effectively reduced performance fluctuations, which are frequent in deep learning models. The application of a non-uniform initial weighting scheme within the AdaBoost framework further enhanced model performance for high Chl-*a* concentration values, which are critical for detecting HABs. The optimization effect of AdaBoost was validated by applying it to data collected from the Ningde area. A robust framework is provided in this study to improve Chl-*a* concentration predictions and offer valuable insights for managing coastal ecosystems facing the challenges of algal blooms.

Key words chlorophyll-*a* concentration forecasting, long short-term memory (LSTM), AdaBoost, deep learning, ensemble learning

Citation Ding Wenxiang, Zhang Caiyun, Li Xueding, Zhang Liyu, Chen Nengwang. 2025. Enhanced forecasting of coastal chlorophyll-*a* through AdaBoost-optimized LSTM models. Acta Oceanologica Sinica, 44(7): 147–160, doi: 10.1007/s13131-025-2492-2

1 Introduction

Harmful algal blooms (HABs) are significant environmental events occurring in coastal waters worldwide. They are driven by nutrient enrichment, hydrodynamic changes, and climatic factors (Moline et al., 2004; Takahashi et al., 2009; Saba et al., 2014). They pose serious threats to marine ecosystems, coastal economies, and human health through degraded water quality, mass mortality of fish, and contamination of seafood with harmful toxins (Tester and Stumpf, 1998; Anderson et al., 2010; Wang et al., 2014; Yu et al., 2018). Effective prediction of HABs is crucial for mitigating their adverse impacts and providing early warnings to enable timely intervention.

Chlorophyll-*a* (Chl-*a*), a proxy for phytoplankton biomass, has become a key indicator for monitoring and forecasting algal bloom events (Strutton et al., 2011; Sarangi, 2012; Harding et al., 2016). Rapid increase in

Foundation item: The National Natural Science Foundation of China under contract No. U22A20579; Guangxi Key R&D Program of China under contract No. GUIKE AB25069453; the Integrated Marine and Fishery Services Program of Fujian Provincial Department of Ocean and Fisheries under contract No. FYZF-YJYB-2025-1-2; the Science and Technology Program of Fujian Province, China, under contract No. 2023Y4001; the Science and Technology Program of Xiamen, China, under contract No. 3502Z20226021.
*Corresponding author, E-mail: cyzhang@xmu.edu.cn

Chl-*a* concentration is often correlated with the initial stages of bloom formation (Siegel et al., 2002), making accurate Chl-*a* concentration forecasting crucial for HAB early warning systems (Vahtera et al., 2007; Nausch et al., 2012; Tian et al., 2019). However, traditional forecasting models, including physical and statistical methods, have limitations in addressing the complex, nonlinear relationships between environmental variables that drive algal blooms (Stumpf et al., 2009; Anderson et al., 2010; Lee and Lee, 2018). These models often fail to predict extreme events, including rapid increases in Chl-*a* levels, which are critical for timely HAB predictions.

In recent years, machine-learning and deep-learning techniques have emerged as powerful tools for overcoming these limitations, offering enhanced capabilities to model complex interactions in environmental data (Tian et al., 2019; Barzegar et al., 2020; Jin et al., 2021). Long short-term memory (LSTM) networks have emerged as a highly effective tool for time series forecasting, particularly in ecological systems where temporal dependencies are critical (Lee and Lee, 2018; Yussof et al., 2021). LSTM networks have been successfully applied in various ecological forecasting scenarios, such as water quality prediction (Gao et al., 2023), hydrological modeling (Cho and Kim, 2022), and HAB forecasting (Liu et al., 2022). They are well suited for predicting Chl-*a* concentration due to their capacity to capture both long-term dependencies and short-term variations in environmental data. These characteristics improve a model's ability to forecast bloom events with greater accuracy.

While LSTM models offer significant advantages, their performance can be further enhanced through ensemble learning techniques. Ensemble learning is a machine learning paradigm that combines multiple base (weak) learners to improve predictive performance by aggregating their outputs (Hu et al., 2021; Zhang et al., 2022). Common ensemble learning methods, such as bagging, stacking, and boosting, can significantly improve the robustness and predictive power of machine-learning models (Alfaro et al., 2013; Talukder et al., 2022; Mohammed and Kora, 2023; Chen et al., 2024).

Adaptive Boosting (AdaBoost), a widely used boosting algorithm, transforms weak learners into strong ones through an iterative process that gives higher weights to misclassified samples. This mechanism allows AdaBoost to focus on difficult-to-predict cases, producing more accurate predictions overall (Jiang et al., 2019). AdaBoost has been successfully applied in various fields, including image recognition and financial forecasting (Wu and Gao, 2018; Hu et al., 2020), but its potential for enhancing deep-learning models in ecological time-series forecasting, particularly in marine ecosystems, remains underexplored. AdaBoost has shown promise in improving model performance in environmental applications including rainfall-runoff prediction (Liu et al., 2014) and air quality forecasting (Liang et al., 2020), yet its integration with LSTM networks for forecasting Chl-*a* concentration in marine ecosystems is a novel approach that has not been thoroughly investigated.

Xiamen Bay is a semi-enclosed bay that is located on the western coast of the Taiwan Strait. It frequently experiences algal blooms as a result of nutrient loading from the Jiulong River (Chen et al., 2013). The unique hydrodynamic conditions and anthropogenic pressures in the bay make it highly susceptible to eutrophication and HABs (Wang et al., 2018; Chen et al., 2021). Accurate forecasts of Chl-*a* concentration in Xiamen Bay are crucial for developing effective management strategies given the significant ecological and economic risks posed by these blooms. Ding et al. (2022) demonstrated that LSTM models could successfully forecast Chl-*a* concentration in Xiamen Bay. However, their study highlighted the need for further optimization, particularly in improving the accuracy of high Chl-*a* concentration predictions indicative of imminent algal bloom events.

In this study, we investigated the application of the AdaBoost algorithm to optimize the performance of LSTM models for Chl-*a* concentration forecasting in Xiamen Bay. By combining multiple weak learners, the AdaBoost-enhanced LSTM model improves Chl-*a* concentration forecasting accuracy, particularly during periods of elevated bloom activity. AdaBoost provides significant enhancement in model performance by focusing on the most challenging cases, resulting in more reliable predictions for both low and high Chl-*a* concentration. The primary objective of this research is to evaluate the effectiveness of AdaBoost in optimizing LSTM-based Chl-*a* concentration forecasts for Xiamen Bay. We hope to provide an improved forecasting tool for mitigating the effects of HABs and supporting the sustainable management of coastal ecosystems by combining deep learning with ensemble learning.

2 Material and methods

2.1 *Data*

Environmental data were collected from a buoy located in Xiamen Bay. The data span the period from January 2008 to June 2021, with a gap from January 2019 to August 2020 due to equipment downtime. Additionally, we collected monitoring data from the Ningde area from September 2017 to July 2018 to validate the model's performance in other regions. The dataset included four key parameters: water temperature, dissolved oxygen (DO), pH, and Chl-*a* concentration. Measurements were recorded every 30 min, resulting in 48 data points per day. These parameters were measured using the YSI EXO2 platform, a water quality monitoring system designed for continuous data collection under dynamic environmental conditions. This dataset was provided by the Xiamen Marine and Fisheries Research Institute. Anomalies in the dataset occurred occasionally due to equipment malfunctions caused by conditions such as wind, waves, currents, or biofouling on the instrument probes. Quality control of the

monitoring data was performed following the methods outlined by Zhang et al. (2009).

Buoy data and daily average meteorological data were collected for the same period from a nearby meteorological station. The meteorological parameters included precipitation, average air pressure, average air temperature, average relative humidity, sunshine hours, minimum air pressure, minimum air temperature, maximum air pressure, maximum air temperature, average two-minute wind speed, and maximum wind speed. These data were obtained from the National Meteorological Information Center (https://data.cma.cn). For the locations of the ecological buoy and the meteorological station please refer to Fig. 1 in Ding et al. (2022). The statistical characteristics of all collected parameters are summarized in Table 1.

For the construction of the LSTM Chl-*a* concentration forecasting model the data were divided by date into training and testing datasets. The training dataset consisted of monitoring data from 2008 to 2018, while the testing dataset included data from September 2020 to June 2021. Prior to input into the LSTM models, all data were normalized using the following equation:

$$x_i' = \frac{x_i - \bar{x}}{S}, \tag{1}$$

where x_i' is the normalized value, x_i is the original parameter value, $\bar{x}$ is the mean of the parameter, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, n is the number of data, and S is the standard deviation, $S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

Previous studies have suggested that evaluating bloom dynamics based on biomass change rates, rather than absolute concentration, can improve forecasting accuracy (Sverdrup, 1953; Behrenfeld and Boss, 2014). Tian et al. (2017, 2019) used the Chl-*a* concentration change rate (ΔChl) in their machine learning model to improve Chl-*a* concentration forecasts and Ding et al. (2022) used the relative Chl-*a* concentration change rate (ΔRChl) in their LSTM model. In this study, we adopted the approach of Ding et al. (2022) and used ΔRChl as the output for the LSTM model to generate forecast results.

2.2 *LSTM model*

LSTM networks, a specialized variant of recurrent neural networks (RNNs), are designed to overcome key limitations of traditional RNNs, particularly the vanishing gradient problem, which makes learning long-term dependencies in time-series data difficult (Hochreiter and Schmidhuber, 1997). LSTMs incorporate memory cells and gating mechanisms—the input, forget, and output gates—that enable them to effectively capture both short-term and long-term temporal dependencies. Differently from standard RNNs, where the influence of earlier inputs fades over time, LSTMs use these gates to control the flow of information, enabling them to retain relevant information across longer sequences. The gating mechanisms are defined as follows:

$$\text{input value: } z = \tanh(\boldsymbol{W}_z[h_{t-1}, x_t]) - z = \tanh(\boldsymbol{W}_z[h_{t-1}, x_t]), \tag{2}$$

$$\text{input gate: } I = \text{sigmoid}(\boldsymbol{W}_I[h_{t-1}, x_t]), \tag{3}$$

$$\text{forgotten gate: } f = \text{sigmoid}(\boldsymbol{W}_f[h_{t-1}, x_t]), \tag{4}$$

$$\text{output gate: } o = \text{sigmoid}(\boldsymbol{W}_o[h_{t-1}, x_t]), \tag{5}$$

Table 1. Statistical summary [mean, standard deviation (SD), minimum (Min), and maximum (Max)] of the observational variables

Parameter	Unit	Mean	SD	Min	Max	*r*
Water temperature	℃	22.8	5.9	11.5	32.2	0.06**
Dissolved oxygen/%	−	97.7	16.1	66.7	199.4	0.57**
pH	−	8.0	0.2	7.5	8.7	0.36**
Chl-*a* concentration	μg/L	3.5	5.6	0.1	91.0	1.00**
Precipitation	mm	3.7	12.2	0	172.7	−0.02
Average air pressure	hPa	997.5	6.5	968.6	1 017.2	−0.06**
Average air temperature	℃	21.6	6.1	3.9	31.9	0.07**
Average relative humidity	−	75.4	12.8	23.0	100.0	0.04*
Sunshine hours	h	5.3	4.1	0	13.0	0.12**
Minimum air pressure	hPa	995.2	6.5	948.9	1 013.2	−0.06**
Minimum air temperature	℃	19.0	6.1	1.1	29.3	0.06**
Maximum air pressure	hPa	999.5	6.6	973.6	1 019.7	−0.06**
Maximum air temperature	℃	25.7	6.3	7.8	38.5	0.08**
Average 2-min wind speed	m/s	2.6	1.0	0.6	9.8	−0.05**
Maximum wind speed	m/s	5.2	1.6	2.1	36.4	0

Note: The correlation coefficient with Chl-*a* concentration (*r*) is also shown. * means $p < 0.05$, and ** means $p < 0.01$. − means no unit.

$$\text{new state: } c_t = f \cdot c_{t-1} + I \cdot z, \tag{6}$$

$$\text{output value: } h_t = o \cdot \tanh c_t, \tag{7}$$

where $\boldsymbol{W}_z[\cdot,\cdot]$, $\boldsymbol{W}_I[\cdot,\cdot]$, $\boldsymbol{W}_f[\cdot,\cdot]$, and $\boldsymbol{W}_o[\cdot,\cdot]$ are four parameter matrices; sigmoid function is defined as sigmoid(a)=1/(1+e^{-a}); x_t is the input at time t; h_{t-1} is the hidden state from the previous time step; z is the candidate cell state at time step t; I, f, and o are the input gate, forget gate, and output gate at time step t, respectively; and c_t is the cell state at time step t.

In this study, we constructed an LSTM model with seven layers, consisting of a sequence input layer, two LSTM layers, two dense (fully connected) layers, a dropout layer, and a regression output layer (Fig. 1). The model outputs ΔRChl, which is then used to calculate the Chl-*a* concentration for the following day based on the current day's concentration. Dropout is a regularization technique that prevents overfitting by randomly dropping units (along with their connections) during training (Murphy et al., 2014; LeCun et al., 2015). To minimize the risk of overfitting we applied dropout between the two LSTM layers. The structure of this LSTM model has demonstrated strong forecasting performance in Chl-*a* concentration predictions in several studies (Ding et al., 2022; Zhang et al., 2024).

2.3 *AdaBoost algorithm*

AdaBoost is an ensemble learning algorithm that combines multiple weak learners to form a strong learner (Freund and Schapire, 1997). It works by iteratively training weak learners, with each new model focusing on the instances that have been misclassified or poorly predicted by the previous models. The predictions from all weak learners are then combined to produce a final model that delivers improved performance. AdaBoost has been widely applied to enhance the accuracy of machine-learning models across various applications (Walker and Jiang, 2019; Sun et al., 2020; Wu et al., 2020).

In this study, we explored how the AdaBoost algorithm affected the LSTM model's performance for Chl-*a* concentration forecasting. The architecture of the AdaBoost-enhanced LSTM model is shown in Fig. 2. The AdaBoost algorithm improves forecasting accuracy by combining multiple LSTM models, each trained on reweighted versions of the data. The LSTM Chl-*a* concentration forecast model served as the base (weak) learner in this study.

A common strategy in forecasting is to average the results of multiple predictions to reduce the influence of random errors. While this approach helps minimize the occurrence of extreme mispredictions, it may also downplay exceptionally accurate predictions. In contrast, AdaBoost improves forecasting accuracy by using a weighted average that prioritizes the predictions made by stronger models. The key feature of the AdaBoost algorithm is its ability to assign weights to each weak learner based on its individual training errors, allowing the model to focus more on difficult cases and improve overall performance.

The ultimate goal of training a machine learning model is to minimize the discrepancies between the forecast results and actual observations. However, minor discrepancies often exist in practice. When implementing the AdaBoost algorithm, it is critical to assess the accuracy of the forecast results based on the training data. We established a threshold for the absolute error of these forecasts. If the absolute error was less than this threshold, the forecast was deemed correct; otherwise, it was considered incorrect.

As shown in Fig. 3, the majority of absolute errors from the weak forecast model's training data results were below 0.05 μg/L, with a probability of 83.8% (34.1% + 49.7%). Even when a model is adequately trained, some forecast results may still exhibit errors greater than 0.05 μg/L. The AdaBoost algorithm is designed to reduce this type of underperformance. Therefore, we set 0.05 μg/L as the threshold. When the absolute error of the training data forecast result was below this threshold, the forecast was considered correct; if it exceeded this value, the forecast was deemed incorrect.

2.4 *Model optimization process*

The detailed optimization process using the AdaBoost

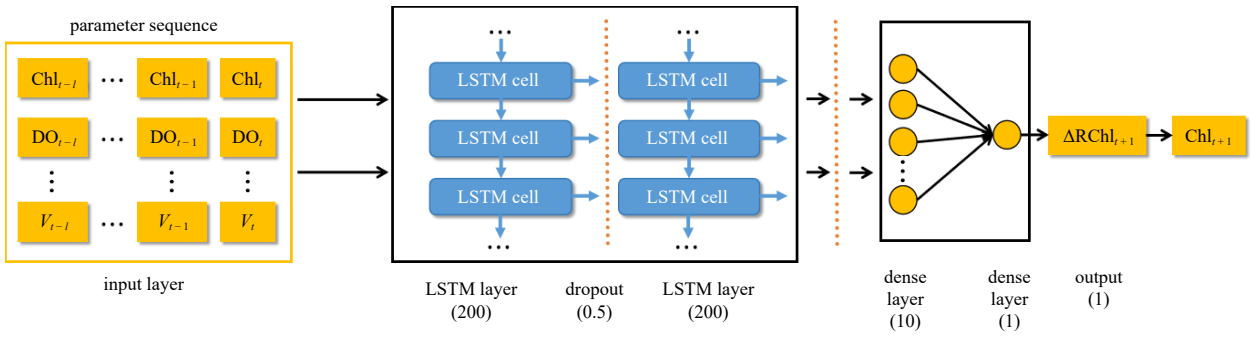

Fig. 1. Architecture of the LSTM model for Chl-*a* concentration prediction. Chl$_t$, DO$_t$, and V_t represent the Chl-*a* concentration, dissolved oxygen, and average 2-min wind speed at time t, respectively. l denotes the length of the input time series. Numbers in parentheses indicate the number of nodes in each layer [e.g., LSTM layer (200) denotes 200 units), except for the dropout layer, where (0.5) represents the dropout rate].

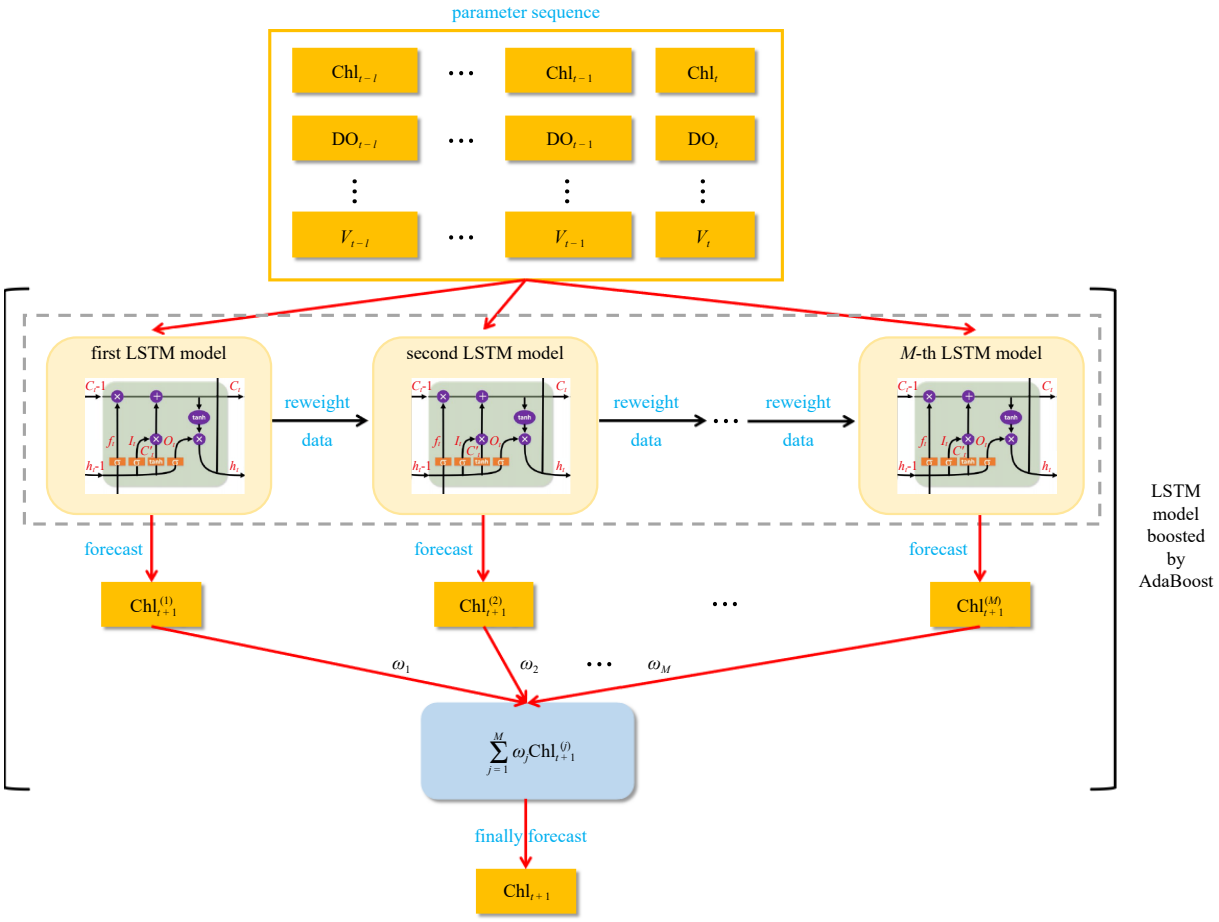

Fig. 2. Architecture of the AdaBoost-enhanced LSTM model for Chl-*a* concentration forecasting. ω_j represents the weight of the *j*-th LSTM model, and $\mathrm{Chl}_{t+1}^{(M)}$ denotes the Chl-*a* concentration at time $t+1$ predicted by the *M*-th LSTM model.

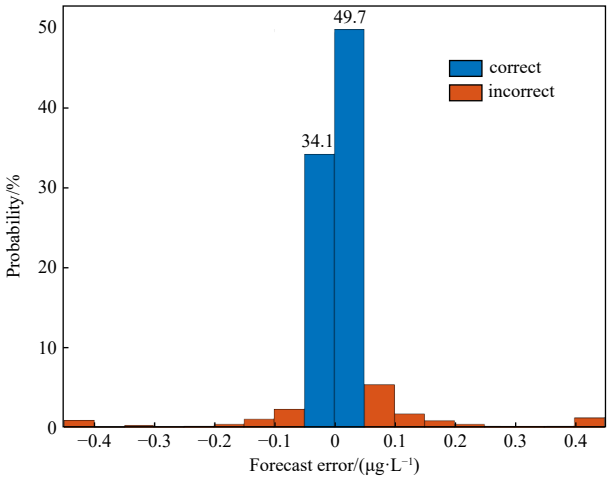

Fig. 3. Distribution of training data forecast absolute errors of the LSTM Chl-*a* concentration forecast model.

algorithm is illustrated in Fig. 4. It consists of the following steps:

Step 1. Training data preprocessing

All training data was standardized and organized into a time series format suitable for input into the LSTM model, as described in Section 2.1.

Step 2. Initialize sample weight

Each training sample was assigned an initial weight. In the traditional AdaBoost algorithm, the initial weights of training samples are typically uniformly distributed,

$$D_1(i) = 1/m, \tag{8}$$

where m is the total number of training samples, and $D_1(i)$ represents the weight of the i-th training sample for the first weak learner.

However, our primary goal was to achieve early warning of algal blooms through the forecast of Chl-*a* concentration. Generally, algal blooms are accompanied by high Chl-*a* levels (Siegel et al., 2002; Tang et al., 2003). Therefore, accurate forecasting of high Chl-*a* values is particularly important. To enhance the model's forecasting ability for high Chl-*a* samples, we used non-uniform initial weight,

$$D_1(i) = \frac{\mathrm{Chl}_i}{\sum_{i=1}^{m} \mathrm{Chl}_i}, \tag{9}$$

where Chl_i represents the Chl-*a* concentration of the i-th

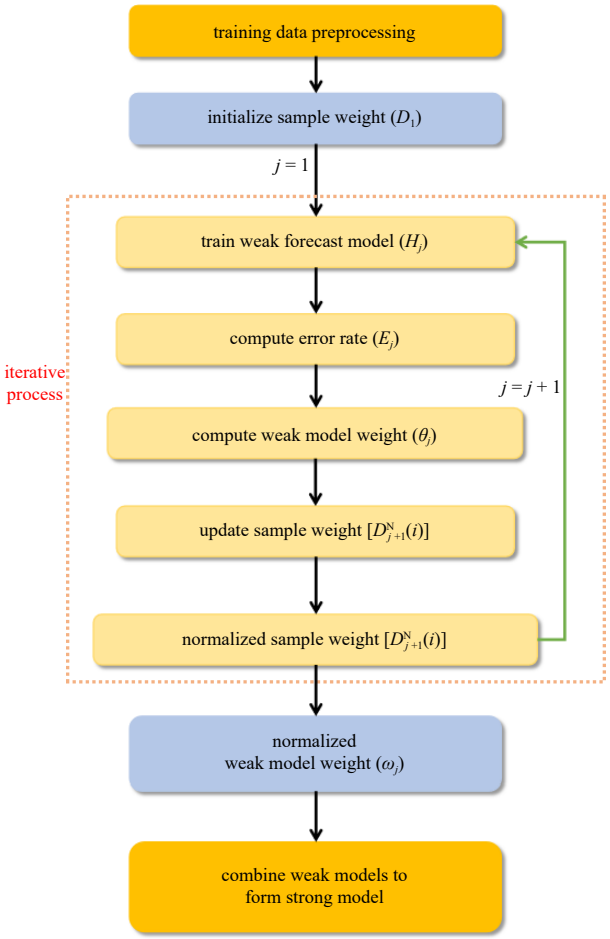

Fig. 4. Flowchart of the AdaBoost algorithm.

training sample.

Step 3. Iterative process

The foundation of the AdaBoost algorithm is an iterative process that repeats for a specified number of iterations k.

For each iteration j = 1, 2, …, k:

Step 3.1. Train weak forecast model

A weak forecast model H_j (a LSTM model) was trained using initial sample weight $D_j(i)$. The goal was to minimize the weighted error.

Step 3.2. Compute error rate

After training, the weighted error E_j of the weak forecast model H_j was calculated,

$$E_j = \sum_{i=1}^{m} \begin{cases} D_j(i), & \left|\mathrm{Chl}_i - O_j(i)\right| \geqslant 0.05, \\ 0, & \left|\mathrm{Chl}_i - O_j(i)\right| < 0.05, \end{cases} \tag{10}$$

where $O_j(i)$ represents the forecast result of the weak forecasting model H_j for the i-th training sample.

Step 3.3. Compute weak model weight

The weight for the weak forecast model H_j was computed,

$$\theta_j = \frac{1}{2\mathrm{e}^{|E_j|}}, \tag{11}$$

where the weight reflects the weak forecast model's performance, with better models receiving higher weights.

Step 3.4. Update sample weight

The weights of the training samples were updated to emphasize those that were incorrectly predicted as follows:

$$D_{j+1}(i) = \begin{cases} a \cdot D_j(i), & \left|\mathrm{Chl}_i - O_j(i)\right| \geqslant 0.05, \\ D_j(i), & \left|\mathrm{Chl}_i - O_j(i)\right| < 0.05, \end{cases} \tag{12}$$

where a is a coefficient slightly greater than 1 (set to 1.1 in this study). Samples that were incorrectly predicted had their weights multiplied by a, making them more influential in the next iteration.

Step 3.5. Normalize sample weight

The updated sample weights were normalized so that they summed to 1,

$$D^{\mathrm{N}}_{j+1}(i) = \frac{D_{j+1}(i)}{\sum_{i=1}^{m} D_{j+1}(i)}. \tag{13}$$

Step 4. Normalize weak model weight

After completing the iterative training process, the weight assigned to each weak forecasting model was collected. To ensure that the combined influence of all weak models was appropriately scaled, these weights were normalized so that their total sum equaled 1,

$$\omega^{\mathrm{N}}_j = \frac{\theta_j}{\sum_{j=1}^{M} \theta_j}. \tag{14}$$

Step 5. Combine weak models to form a strong model

After completing all iterations, the final strong forecast model was constructed by aggregating the predictions of all weak forecasting models, each weighted by their respective normalized weights ω^{N}_j,

$$\mathrm{Chl}_{t+1} = \sum_{j=1}^{M} \omega_j \cdot \mathrm{Chl}^{(j)}_{t+1}, \tag{15}$$

where $\mathrm{Chl}^{(j)}_{t+1}$ represents the forecasted Chl-a concentration by the j-th weak forecast at time t + 1; Chl_{t+1} denotes forecast Chl-a concentration by the strong forecast model at time t + 1; and M is the total number of weak forecasting models. In this study, M was set to 10 to balance optimization performance and training time. When M exceeded 10, there was no significant improvement in forecasting accuracy, while the model's training time increased. This weighted aggregation leverages the collective strengths of all weak models, enhancing the overall

predictive accuracy and robustness of the strong forecasting model.

2.5 *Model evaluation metrics*

The model's performance was evaluated using three metrics: the correlation coefficient (r), root mean square error (RMSE), and absolute error. The r assesses the relationship between predicted and observed results, with values closer to 1 indicating higher accuracy. The RMSE and absolute error quantify the differences between predicted and observed values, with smaller values reflecting better model performance.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_\mathrm{s}}\sum_{i=1}^{n_\mathrm{s}}(Y_i - y_i)^2}, \tag{16}$$

$$r = \frac{\sum_{i=1}^{n_\mathrm{s}}(Y_i - \overline{Y}_i)(y_i - \overline{y}_i)}{\sqrt{\sum_{i=1}^{n_\mathrm{s}}(Y_i - \overline{Y}_i)^2 \sum_{i=1}^{n_\mathrm{s}}(y_i - \overline{y}_i)^2}}, \tag{17}$$

$$\text{absolute error}_i = |Y_i - y_i|, \tag{18}$$

where Y_i and y_i represent the observed and predicted values, respectively, $\overline{Y}_i$ and $\overline{y}_i$ are their mean values, and n_s is the number of samples.

3 Results

To minimize the potential influence of random errors, we conducted ten independent tests. Each test involved training ten weak forecasting models and one strong forecasting model, the latter generated by combining the weak models using the AdaBoost algorithm. The performance of these models was evaluated using the correlation coefficient (r) and RMSE, as summarized in Tables 2 and 3, respectively.

The results in Table 2 demonstrate that the AdaBoost-optimized strong forecasting models consistently outperformed the individual weak forecasting models. In every test, r of the strong models exceeded the average r of the weak models. Moreover, in most tests, r of the strong models were higher than those of all individual weak models. For example, in Test 1, the strong model achieved

Table 2. Correlation coefficient (r) of the Chl-*a* concentration forecast results for the weak forecast models and the AdaBoost-optimized strong forecast model

Forecast model	r									
	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
Weak model 1	0.910	0.889	0.928	0.924	0.918	0.912	0.915	0.924	0.918	0.906
Weak model 2	0.934	0.901	0.921	0.947	0.916	0.922	0.922	0.900	0.922	0.926
Weak model 3	0.933	0.943	0.916	0.935	0.936	0.909	0.911	0.911	0.916	0.936
Weak model 4	0.935	0.885	0.901	0.911	0.915	0.926	0.920	0.937	0.917	0.915
Weak model 5	0.907	0.910	0.924	0.893	0.903	0.905	0.913	0.933	0.908	0.930
Weak model 6	0.934	0.925	0.925	0.933	0.927	0.928	0.922	0.921	0.883	0.914
Weak model 7	0.914	0.909	0.934	0.914	0.911	0.934	0.921	0.901	0.925	0.940
Weak model 8	0.916	0.937	0.915	0.948	0.905	0.922	0.932	0.912	0.909	0.927
Weak model 9	0.910	0.923	0.917	0.941	0.903	0.929	0.904	0.916	0.890	0.917
Weak model 10	0.892	0.939	0.923	0.915	0.949	0.894	0.901	0.912	0.926	0.922
Weak model (average)	0.919	0.916	0.921	0.926	0.918	0.918	0.916	0.917	0.911	0.923
Strong model	0.936	0.936	0.940	0.944	0.939	0.936	0.935	0.934	0.929	0.943

Table 3. RMSE of the Chl-*a* concentration forecast results of the weak forecast models and the AdaBoost-optimized strong forecast model

Forecast model	RMSE/(μg·L^{-1})									
	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
Weak model 1	1.243	1.334	1.077	1.105	1.125	1.190	1.167	1.120	1.158	1.202
Weak model 2	1.023	1.263	1.214	0.937	1.211	1.115	1.149	1.270	1.176	1.117
Weak model 3	1.024	0.961	1.143	1.011	1.013	1.236	1.255	1.279	1.219	1.003
Weak model 4	1.014	1.376	1.252	1.226	1.185	1.105	1.113	1.001	1.171	1.145
Weak model 5	1.243	1.190	1.118	1.322	1.293	1.226	1.243	1.078	1.194	1.047
Weak model 6	1.023	1.097	1.098	1.023	1.060	1.057	1.204	1.133	1.424	1.170
Weak model 7	1.150	1.239	1.018	1.167	1.217	1.025	1.173	1.321	1.120	0.993
Weak model 8	1.187	1.014	1.156	0.902	1.256	1.097	1.052	1.186	1.309	1.079
Weak model 9	1.182	1.144	1.153	1.020	1.316	1.090	1.248	1.190	1.365	1.148
Weak model 10	1.317	1.034	1.115	1.150	0.907	1.352	1.285	1.173	1.075	1.194
Weak model (average)	1.141	1.165	1.134	1.086	1.158	1.149	1.189	1.175	1.221	1.110
Strong model	1.000	1.014	0.970	0.937	0.987	1.004	1.037	1.039	1.078	0.944

an r of 0.936, which surpassed both the average r and r of all ten weak models. In Tests 2, 4, 5, and 8, although r of the strong model was slightly lower than that of up to three weak models, it still exceeded the average performance.

Similarly, Table 3 indicates that the RMSE of the strong forecasting models was consistently lower than the average RMSE of the weak models. In most tests, the strong models achieved lower RMSE values than each individual weak model, underscoring the effectiveness of AdaBoost in reducing prediction errors. Even in instances where some weak models had slightly lower RMSE than the strong model, typically only one weak model per test outperformed the strong model.

The time series of forecasted Chl-*a* concentration from September 2020 to June 2021 further illustrates the performance improvement (Fig. 5). The strong forecasting model produced more stable and accurate predictions than the individual weak models. Average r across the weak models was 0.918, whereas the strong model achieved an r of 0.937. This enhancement highlights the AdaBoost algorithm's ability to improve model performance by compensating for individual model errors and addressing weaknesses.

The distribution of absolute error (Fig. 6) also demonstrates that the strong model significantly improved the error profile. The distribution of absolute errors below 0.25 μg/L increased from 44% in the weak models to 49% in the strong model. Furthermore, the probability of errors exceeding 5.00 μg/L was halved in the strong model, indicating its enhanced capability to reduce large prediction errors. This ability is critical for forecasting extreme events such as HABs.

Variability in model performance was further assessed using boxplots of r and RMSE (Fig. 7). The weak models exhibited greater variability, reflecting inconsistencies in their predictions due to factors such as random parameter initialization. In contrast, the strong model demonstrated reduced variability, with higher r and lower RMSE, confirming that the AdaBoost effectively stabilized model performance.

4 Discussion

4.1 *Optimization performance of AdaBoost*

The AdaBoost algorithm is renowned for enhancing the performance of weak learners by combining them into a strong ensemble model (Freund and Schapire, 1997). Its application has been widespread in various domains, including ecological modeling (Kadavi et al., 2018; Peng et al., 2020), and in predicting HABs and Chl-*a* concentration (Aláez et al., 2021; Shin et al., 2021). In this study, the AdaBoost algorithm was used to optimize LSTM

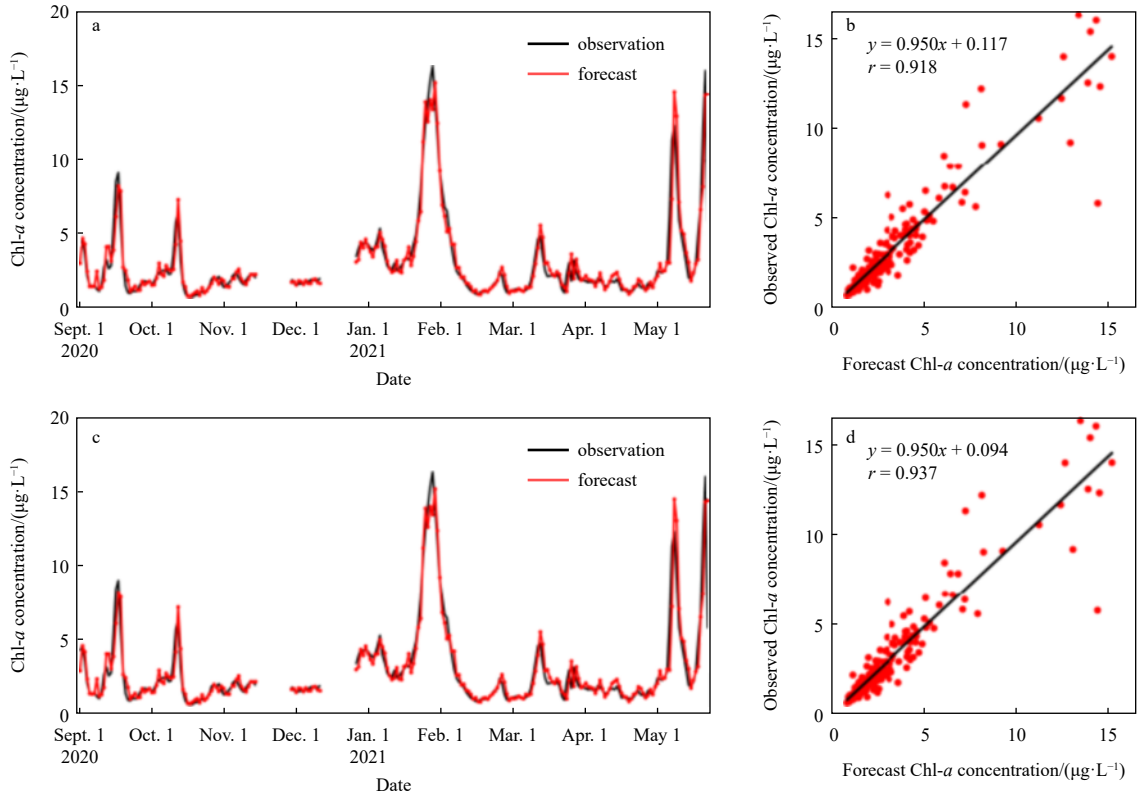

Fig. 5. Time series of observed and predicted Chl-*a* concentration using individual weak forecasting models (a) and the AdaBoost-optimized strong forecasting model (c) from September 2020 to June 2021, and scatter plots showing the correlations between observed and predicted values corresponding to a (b) and c (d), respectively. The black solid lines in b and d represent linear regression fits between the observed and predicted values, respectively.

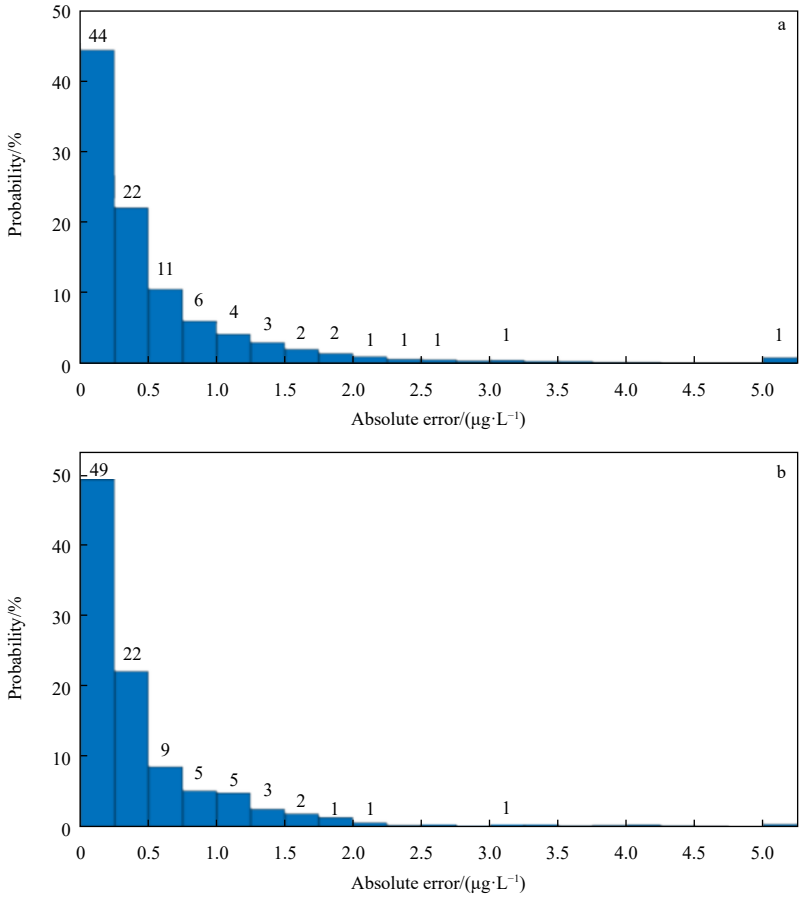

Fig. 6. Distribution of absolute errors in Chl-*a* concentration forecasts for individual weak models (a) and the AdaBoost-optimized strong model (b).

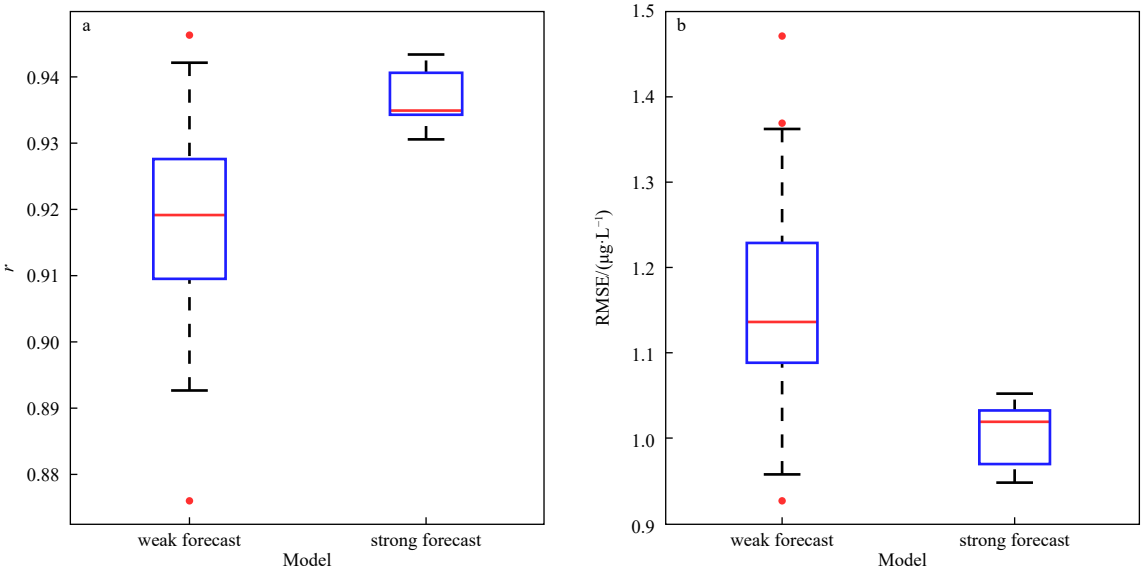

Fig. 7. Boxplots of *r* (a) and RMSE (b) between the forecast and observed Chl-*a* concentration for the weak forecasting models and the AdaBoost-optimized strong forecasting model. The boxplots use the following conventions: the blue boxes represent the interquartile range (IQR) between the first quartile (Q1) and the third quartile (Q3); the red horizontal line inside each box indicates the median value; the black dashed lines (whiskers) extend to the minimum and maximum values within 1.5 times the IQR from Q1 and Q3, respectively; and the red dots denote outliers beyond the whisker range.

models for Chl-*a* concentration forecasting in Xiamen Bay. The results showed that the AdaBoost-optimized strong models consistently outperformed individual weak models (Tables 2 and 3).

AdaBoost enhances model performance by iteratively focusing on the errors of previous models. Each subsequent weak model is trained to correct the shortcomings of its predecessors, emphasizing samples that were previously mispredicted. This iterative refinement leads to enhanced understanding of the training data and improved prediction accuracy (Walker and Jiang, 2019; Sun et al., 2020).

Our findings verify the effectiveness of AdaBoost in reducing prediction errors and enhancing correlation with observed values. The strong model both improved accuracy within the 0.25 μg/L error range and also significantly reduced large errors exceeding 5.00 μg/L. This is particularly important for forecasting HABs, which are often associated with sudden spikes in Chl-*a* concentration (Anderson et al., 2010; Siegel et al., 2002). Minimizing large prediction errors is vital for timely and reliable early warning of events.

The reduced variability in model performance (Fig. 7) highlights AdaBoost's ability to stabilize deep learning models, which are known to be sensitive to random initialization and training conditions (Goodfellow et al., 2016). AdaBoost reduces the risk of poor predictions by combining multiple weak learners, resulting in more consistent and reliable forecasts.

4.2 *Role of non-uniform initial weight*

In our use of the AdaBoost framework, we incorporated a non-uniform initial weighting scheme to prioritize the accurate prediction of high Chl-*a* concentration associated with HABs. Accurately forecasting these high values is critical for early detection and management of HABs (Siegel et al., 2002; Tang et al., 2003). By assigning higher initial weights to samples with elevated Chl-*a* concentration, the algorithm directed more focus toward these critical cases during training.

Our results demonstrated that the non-uniform weighting scheme led to notable improvements in the model's performance for high Chl-*a* concentration compared to the uniform weighting approach. Specifically, correlation coefficient (r) for high Chl-*a* concentration was higher by 1.5%, and the RMSE was lower by 2.9% (Table 4). In contrast, the improvement for low Chl-*a* concentration was less pronounced, with r only 0.2% higher and RMSE only 1.2% lower. This indicates that the non-uniform weighting scheme was most effective at enhancing the model's ability to predict high Chl-*a* concentration, which are often indicative of bloom events (Fig. 8).

The improved performance for high Chl-*a* concentration can be attributed to the model's increased emphasis on these samples during training. This focus allowed the model to better capture the dynamics of algal bloom formation and intensification, thereby improving the accuracy of early warning systems.

4.3 *Implications for ecological forecasting*

To evaluate the broader applicability of the proposed method, we extended the analysis to the Ningde area, located in the northern part of the Taiwan Strait. Using monitoring data from September 2017 to July 2018, we observed that, even in a different geographical setting, the AdaBoost optimization significantly improved the model's performance, with a notable increase in r and a considerable reduction in RMSE (Fig. 9). These results indicate that the AdaBoost-optimized LSTM model is applicable to other regions, further demonstrating its flexibility and robustness.

The combination of AdaBoost and LSTM in this study is driven by the complementary strengths of both techniques. LSTM, a recurrent neural network, is well suited for capturing temporal dependencies in time-series data, including the dynamics of Chl-*a* concentration (Ding et al., 2022; Zhang et al., 2024). However, LSTM still has limitations in forecasting Chl-*a* concentration, particularly in improving the prediction accuracy for high Chl-*a* concentration (Ding et al., 2022). To address these limitations, we incorporated AdaBoost, an ensemble learning algorithm, which improves model performance by iteratively adjusting weights on misclassified samples and reducing bias. The AdaBoost algorithm helps improve the stability and accuracy of the LSTM model, particularly by focusing on challenging instances that may be crucial for detecting extreme events, such as HABs. This combination allows for both effective temporal modeling and improved robustness, making it an ideal strategy for ecological forecasting. The success of the AdaBoost-optimized LSTM model in this study has significant implications for ecological forecasting. We have demonstrated that ensemble learning methods like AdaBoost can effectively enhance the predictive capabilities of models dealing with complex ecological phenomena, such as HABs, which have substantial environmental and socio-economic impacts (Anderson et al., 2010).

By improving the accuracy of Chl-*a* concentration fore-

Table 4. Comparison of the optimization performance of the AdaBoost algorithm with non-uniform and uniform initial weights for all, high (≥5 μg/L), and low (<5 μg/L) Chl-*a* concentration values

Type of weight	r			RMSE/(μg·L^{-1})		
	All	High	Low	All	High	Low
Non-uniform initial weight	0.937 ± 0.004	1.001 ± 0.044	0.783 ± 0.019	2.487 ± 0.137	0.895 ± 0.005	0.467 ± 0.011
Uniform initial weight	0.934 ± 0.002	1.027 ± 0.022	0.771 ± 0.011	2.562 ± 0.061	0.893 ± 0.003	0.476 ± 0.008

Note: Evaluation metrics values are in the format of mean ± SD.

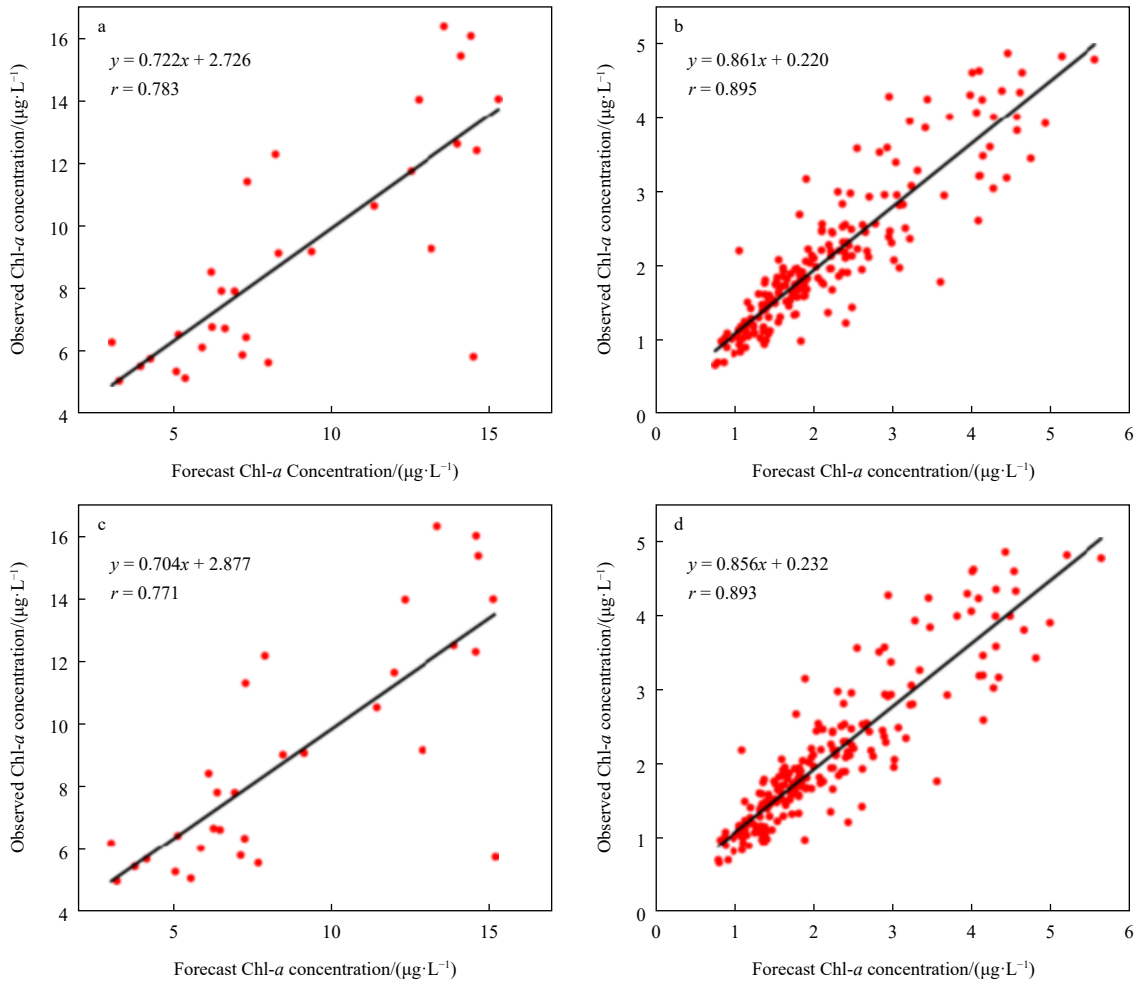

Fig. 8. Scatter plots of observed versus predicted Chl-*a* concentration using the AdaBoost algorithm with non-uniform initial weights (a, b) and uniform initial weights (c, d) for high (⩾5 μg/L) and low (<5 μg/L) Chl-*a* concentration. The black solid lines represent the linear regression fits between the observed and predicted values.

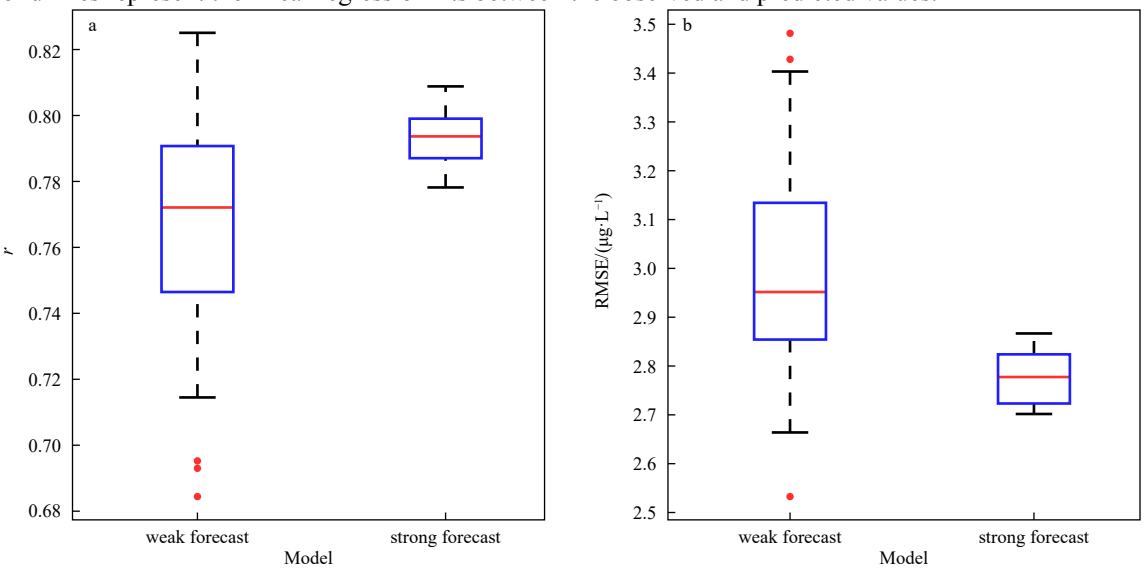

Fig. 9. Boxplots of *r* (a) and RMSE (b) between the forecast and observed Chl-*a* concentration for the weak forecasting models and the AdaBoost-optimized strong forecasting model applied to the Ningde area. The boxplots use the following conventions: the blue boxes represent the interquartile range (IQR) between the first quartile (Q1) and the third quartile (Q3); the red horizontal line inside each box indicates the median value; the black dashed lines (whiskers) extend to the minimum and maximum values within 1.5 times the IQR from Q1 and Q3, respectively; and the red dots denote outliers beyond the whisker range.

casts, particularly for high concentration, the AdaBoost-optimized model provides a valuable tool for early warning systems. Its use will allow coastal managers to make timely interventions, such as issuing public health advisories or implementing mitigation strategies, thereby reducing the adverse effects of HABs on marine ecosystems and human health.

Furthermore, this methodology could be applied to forecasting other ecological variables of interest, such as dissolved oxygen levels, nutrient concentrations, or harmful toxin levels. The adaptability of AdaBoost in focusing on critical thresholds makes it a versatile tool for ecological forecasting.

4.4 *Limitations and future research*

Despite these promising results, several limitations of the AdaBoost-optimized LSTM model should be acknowledged. One primary limitation is the increased computational cost associated with training multiple weak learners. The ensemble approach requires more computational resources and time than single-model approaches. This may create challenges for real-time forecasting applications without the necessary infrastructure.

Additionally, while the LSTM served as an effective base learner in this study, exploring other machine learning algorithms or hybrid models could further enhance predictive performance. Combining LSTM with models like random forests or gradient boosting machines may leverage complementary strengths and improve robustness (Chen and Guestrin, 2016).

Future research should also consider incorporating additional environmental variables, such as nutrient levels, light availability, and hydrodynamic conditions, to capture a broader range of factors influencing algal blooms. Integrating remote sensing data or developing data assimilation techniques could further improve forecast accuracy and spatial resolution (Stumpf et al., 2009; Tian et al., 2017).

Building on the promising results of our model for Xiamen Bay, we applied it to the Ningde area, demonstrating its potential for use in other regions with similar conditions. However, as both Xiamen Bay and the Ningde area are located on the western side of the Taiwan Strait, further validation is needed to assess the model's applicability to a wider range of marine environments with different ecological characteristics. Additionally, our approach relies on over a decade of data, which may not be available for many coastal areas. The performance of the model for regions with limited data requires further investigation. While the results are promising, the proposed model's reliance on long-term data is limiting, as performance may fluctuate for regions with limited data. Future research should focus on refining the model to enhance its robustness when applied to these areas and exploring its integration with larger-scale ecological monitoring systems to improve its applicability and forecasting capabilities across diverse marine ecosystems.

In conclusion, the AdaBoost-optimized LSTM model demonstrated significant potential for enhancing early warning systems for HABs and offered valuable insights for broader ecological forecasting applications. Continued refinement of these models through the integration of diverse data sources and advanced machine learning techniques will further enhance their predictive capabilities.

5 Conclusions

This study demonstrated that optimizing LSTM models with the AdaBoost algorithm significantly enhanced the accuracy of Chl-*a* concentration forecasting in Xiamen Bay. The AdaBoost-optimized strong forecasting model consistently outperformed individual weak models. Specifically, after optimization, the frequency of predictions with low absolute errors (<0.25 μg/L) increased, while the occurrence of large absolute errors (>5.00 μg/L) decreased. This indicates that AdaBoost not only improved overall prediction accuracy but also effectively reduced large errors. This ability is crucial for detecting extreme events like HABs.

Moreover, the iterative compensation mechanism of the AdaBoost algorithm significantly reduced the performance variability of the model, making it more stable and reliable for complex ecological forecasting tasks than the model would be without the AdaBoost. By introducing non-uniform initial weights within the AdaBoost framework—assigning higher weights to data with high Chl-*a* concentration—the model's ability to learn and capture the dynamic changes of high Chl-*a* concentration was further enhanced. Compared to the traditional uniform weighting method, the non-uniform weighting approach showed a distinct advantage in improving predictions of high Chl-*a* concentration critical for HAB detection.

In conclusion, integrating AdaBoost with LSTM models is an effective method for improving early warning systems for HABs. The enhanced predictive performance, especially with respect to high Chl-*a* concentration, offers valuable insights for other ecological forecasting applications. This study highlights the potential of ensemble learning techniques in advancing environmental monitoring and management, creating the foundation for more reliable and accurate ecological forecasts.

References

Aláez F M B, Palenzuela J M T, Spyrakos E, et al. 2021. Machine learning methods applied to the prediction of *Pseudo-nitzschia* spp. blooms in the Galician *Rias Baixas* (NW Spain). ISPRS International Journal of Geo-Information, 10(4): 199, doi: 10.3390/ijgi10040199

Alfaro E, Gamez M, García N. 2013. adabag: an R package for classification with boosting and bagging. Journal of Statistical Software, 54(2): 1–35

Anderson C R, Sapiano M R P, Prasad M B K, et al. 2010. Predicting potentially toxigenic *Pseudo-nitzschia* blooms in the Chesapeake Bay. Journal of Marine Systems, 83(3–4): 127–140, doi: 10.1016/j.jmarsys.2010.04.003

Barzegar R, Aalami M T, Adamowski J. 2020. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. Stochastic Environmental Research and Risk Assessment, 34(2): 415–433, doi: 10.1007/s00477-020-01776-2

Behrenfeld M J, Boss E S. 2014. Resurrecting the ecological underpinnings of ocean plankton blooms. Annual Review of Marine Science, 6: 167–194, doi: 10.1146/annurev-marine-052913-021325

Chen Pengyi, Chen Zhongbiao, Sun Runxia, et al. 2024. An ensemble learning method to retrieve sea ice roughness from Sentinel-1 SAR images. Acta Oceanologica Sinica, 43(5): 78–90, doi: 10.1007/s13131-023-2248-9

Chen Tianqi, Guestrin C. 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 785–794, doi: 10.1145/2939672.2939785

Chen Baohong, Ji Weidong, Chen Jinmin, et al. 2013. Characteristics of nutrients in the Jiulong River and its impact on Xiamen Water, China. Chinese Journal of Oceanology and Limnology, 31(5): 1055–1063, doi: 10.1007/s00343-013-2263-3

Chen Baohong, Wang Kang, Dong Xu, et al. 2021. Long-term changes in red tide outbreaks in Xiamen Bay in China from 1986 to 2017. Estuarine, Coastal and Shelf Science, 249: 107095, doi: 10.1016/j.ecss.2020.107095

Cho K, Kim Y. 2022. Improving streamflow prediction in the WRF-Hydro model with LSTM networks. Journal of Hydrology, 605: 127297, doi: 10.1016/j.jhydrol.2021.127297

Ding Wenxiang, Zhang Caiyun, Shang Shaoping, et al. 2022. Optimization of deep learning model for coastal chlorophyll *a* dynamic forecast. Ecological Modelling, 467: 109913, doi: 10.1016/j.ecolmodel.2022.109913

Freund Y, Schapire R E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1): 119–139, doi: 10.1006/jcss.1997.1504

Gao Zhenyu, Chen Jinyue, Wang Guoqiang, et al. 2023. A novel multivariate time series prediction of crucial water quality parameters with LSTM networks. Journal of Contaminant Hydrology, 259: 104262, doi: 10.1016/j.jconhyd.2023.104262

Goodfellow I, Bengio Y, Courville A. 2016. Deep Learning. Cambridge: MIT Press

Harding L W Jr, Mallonee M E, Perry E S, et al. 2016. Variable climatic conditions dominate recent phytoplankton dynamics in Chesapeake Bay. Scientific Reports, 6: 23773, doi: 10.1038/srep23773

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

Hu Yi, Qu Boyang, Liang Jing, et al. 2021. A survey on evolutionary ensemble learning algorithm. Chinese Journal of Intelligent Science and Technology (in Chinese), 3(1): 18–33, doi: 10.11959/j.issn.2096-6652.202103

Hu Gensheng, Yin Cunjun, Wan Mingzhu, et al. 2020. Recognition of diseased Pinus trees in UAV images using deep learning and AdaBoost classifier. Biosystems Engineering, 194: 138–151, doi: 10.1016/j.biosystemseng.2020.03.021

Jiang He, Zheng Weihua, Luo Liangqing, et al. 2019. A two-stage minimax concave penalty based method in pruned AdaBoost ensemble. Applied Soft Computing, 83: 105674, doi: 10.1016/j.asoc.2019.105674

Jin D, Lee E, Kwon K, et al. 2021. A deep learning model using satellite ocean color and hydrodynamic model to estimate chlorophyll-*a* concentration. Remote Sensing, 13(10): 2003, doi: 10.3390/rs13102003

Kadavi P R, Lee C W, Lee S. 2018. Application of ensemble-based machine learning models to landslide susceptibility mapping. Remote Sensing, 10(8): 1252, doi: 10.3390/rs10081252

LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature, 521(7553): 436–444, doi: 10.1038/nature14539

Lee S, Lee D. 2018. Improved prediction of harmful algal blooms in four major South Korea's rivers using deep learning models. International Journal of Environmental Research and Public Health, 15(7): 1322, doi: 10.3390/ijerph15071322

Liang Yun-Chia, Maimury Y, Chen Angela Hsiang-Ling, et al. 2020. Machine learning-based prediction of air quality. Applied Sciences, 10(24): 9151, doi: 10.3390/app10249151

Liu Muyuan, He Junyu, Huang Yuzhou, et al. 2022. Algal bloom forecasting with time-frequency analysis: a hybrid deep learning approach. Water Research, 219: 118591, doi: 10.1016/j.watres.2022.118591

Liu Shuang, Xu Jingwen, Zhao Junfang, et al. 2014. Efficiency enhancement of a process-based rainfall–runoff model using a new modified AdaBoost. RT technique. Applied Soft Computing, 23: 521–529, doi: 10.1016/j.asoc.2014.05.033

Mohammed A, Kora R. 2023. A comprehensive review on ensemble deep learning: opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 35(2): 757–774, doi: 10.1016/j.jksuci.2023.01.014

Moline M A, Claustre H, Frazer T K, et al. 2004. Alteration of the food web along the Antarctic Peninsula in response to a regional warming trend. Global Change Biology, 10(12): 1973–1980, doi: 10.1111/j.1365-2486.2004.00825.x

Murphy K, Schölkopf B, Srivastava N, et al. 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1): 1929–1958

Nausch M, Nausch G, Mohrholz V, et al. 2012. Is growth of filamentous cyanobacteria supported by phosphate uptake below the thermocline?. Estuarine, Coastal and Shelf Science, 99: 50–60, doi: 10.1016/j.ecss.2011.12.011

Peng Liheng, Liu Kai, Cao Jingjing, et al. 2020. Combining GF-2 and RapidEye satellite data for mapping mangrove

species using ensemble machine-learning methods. International Journal of Remote Sensing, 41(3): 813–838, doi: 10.1080/01431161.2019.1648907

Saba G K, Fraser W R, Saba V S, et al. 2014. Winter and spring controls on the summer food web of the coastal West Antarctic Peninsula. Nature Communications, 5: 4318, doi: 10.1038/ncomms5318

Sarangi R K. 2012. Observation of algal bloom in the Northwest Arabian sea using multisensor remote sensing satellite data. Marine Geodesy, 35(2): 158–174, doi: 10.1080/01490419.2011.637848

Shin J, Yoon S, Kim Y, et al. 2021. Effects of class imbalance on resampling and ensemble learning for improved prediction of cyanobacteria blooms. Ecological Informatics, 61: 101202, doi: 10.1016/j.ecoinf.2020.101202

Siegel D A, Doney S C, Yoder J A. 2002. The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. Science, 296(5568): 730–733, doi: 10.1126/science.1069174

Strutton P G, Martz T R, DeGrandpre M D, et al. 2011. Bio-optical observations of the 2004 Labrador Sea phytoplankton bloom. Journal of Geophysical Research: Oceans, 116(C11): C11037, doi: 10.1029/2010JC006872

Stumpf R P, Tomlinson M C, Calkins J A, et al. 2009. Skill assessment for an operational algal bloom forecast system. Journal of Marine Systems, 76(1–2): 151–161, doi: 10.1016/j.jmarsys.2008.05.016

Sun Jie, Li Hui, Hamido F, et al. 2020. Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. Information Fusion, 54: 128–144, doi: 10.1016/j.inffus.2019.07.006

Sverdrup H U. 1953. On conditions for the vernal blooming of phytoplankton. Journal du Conseil, 18(3): 287–295, doi: 10.1093/icesjms/18.3.287

Takahashi T, Sutherland S C, Wanninkhof R, et al. 2009. Climatological mean and decadal change in surface ocean pCO_2, and net sea–air CO_2 flux over the global oceans. Deep Sea Research Part Ⅱ: Topical Studies in Oceanography, 56(8–10): 554–577, doi: 10.1016/j.dsr2.2008.12.009

Talukder A, Islam M, Uddin A, et al. 2022. Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. Expert Systems with Applications, 205: 117695, doi: 10.1016/j.eswa.2022.117695

Tang Danling, Kester D R, Ni I H, et al. 2003. In situ and satellite observations of a harmful algal bloom and water condition at the Pearl River estuary in late autumn 1998. Harmful Algae, 2(2): 89–99, doi: 10.1016/S1568-9883(03)00021-0

Tester P A, Stumpf R P. 1998. Phytoplankton blooms and remote sensing: what is the potential for early warning. Journal of Shellfish Research, 17(5): 1469–1471

Tian Wenchong, Liao Zhenliang, Wang Xuan. 2019. Transfer learning for neural network model in chlorophyll-a dynamics prediction. Environmental Science and Pollution Research, 26(29): 29857–29871, doi: 10.1007/s11356-019-06156-0

Tian Wenchong, Liao Zhenliang, Zhang Jin. 2017. An optimization of artificial neural network model for predicting chlorophyll dynamics. Ecological Modelling, 364: 42–52, doi: 10.1016/j.ecolmodel.2017.09.013

Vahtera E, Laamanen M, Rintala J M. 2007. Use of different phosphorus sources by the bloom-forming cyanobacteria *Aphanizomenon flos-aquae* and *Nodularia spumigena*. Aquatic Microbial Ecology, 46: 225–237, doi: 10.3354/ame046225

Walker K W, Jiang Zhehan. 2019. Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: a machine-learning approach. The Journal of Academic Librarianship, 45(3): 203–212, doi: 10.1016/j.acalib.2019.02.013

Wang Baodong, Xin Ming, Wei Qinsheng, et al. 2018. A historical overview of coastal eutrophication in the China Seas. Marine Pollution Bulletin, 136: 394–400, doi: 10.1016/j.marpolbul.2018.09.044

Wang Qing, Zhu Liangsheng, Wang Dongxiao. 2014. A numerical model study on multi-species harmful algal blooms coupled with background ecological fields. Acta Oceanologica Sinica, 33(8): 95–105, doi: 10.1007/s13131-014-0459-9

Wu Yungao, Gao Jianwei. 2018. AdaBoost-based long short-term memory ensemble learning approach for financial time series forecasting. Current Science, 115(1): 159–165, doi: 10.18520/cs/v115/i1/159-165

Wu Yanli, Ke Yutian, Chen Zhuo, et al. 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. CATENA, 187: 104396, doi: 10.1016/j.catena.2019.104396

Yu Rencheng, Lü Songhui, Liang Yubo. 2018. Harmful algal blooms in the coastal waters of China. In: Glibert P M, Berdalet E, Burford M A, et al. Global Ecology and Oceanography of Harmful Algal Blooms. Cham: Springer, 309–316.

Yussof F N, Maan N, Reba M N M. 2021. LSTM networks to improve the prediction of harmful algal blooms in the west coast of Sabah. International Journal of Environmental Research and Public Health, 18(14): 7650, doi: 10.3390/ijerph18147650

Zhang Caiyun, Ding Wenxiang, Zhang Liyu. 2024. Impacts of missing buoy data on LSTM-based coastal chlorophyll-a forecasting. Water, 16(21): 3046, doi: 10.3390/w16213046

Zhang Caiyun, Zhang Xuemin, Shang Shaoling. 2009. Survey of ensemble classification methods for complex data stream. Study on quality control of automoy data in Xiamen west area. In: Chinese Society For Environmental Sciences, ed. Proceedings of the 2009 Annual Conference (in Chinese). Beijing: Beihang University Press, 582–586

Zhang Xilong, Han Meng, Chen Zhiqiang, et al. 2022. Survey of ensemble classification methods for complex data stream. Journal of Guangxi Normal University (Natural Science Edition) (in Chinese), 40(4): 1–21, doi: 10.16088/j.issn.1001-6600.2021071102