Water Research 219 (2022) 118591

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Algal bloom forecasting with time-frequency analysis: A hybrid deep learning approach

Muyuan Liu [a], Junyu He [a,b], Yuzhou Huang [a], Tao Tang [a], Jing Hu [a], Xi Xiao [a,c,*]

[a] Ocean College, Zhejiang University, #1 Zheda Road, Zhoushan, Zhejiang 316021, China
[b] Ocean Academy, Zhejiang University, #1 Zheda Road, Zhoushan, Zhejiang 316021, China
[c] Key Laboratory of Watershed Non-point Source Pollution Control and Water Eco-security of Ministry of Water Resources, College of Environmental and Resources Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

ARTICLE INFO

Keywords:
Algal bloom forecasting
Deep learning
Long-short-term-memory (LSTM)
Wavelet analysis

ABSTRACT

The rapid emergence of deep learning long-short-term-memory (LSTM) technique presents a promising solution to algal bloom forecasting. However, the discontinuous and non-stationary processes within algal dynamics still largely limit the functions of LSTMs. To overcome this challenge, an advanced time-frequency wavelet analysis (WA) technique was introduced to enhance the prediction accuracy of LSTMs. Herein, the novel hybrid approach (named WLSTM) successfully decreased the algal forecasting inaccuracy of classic LSTMs by 41% ± 8% in Lake Mendota (Wisconsin, USA), with powerful one-step-ahead predictions at hourly, daily, and monthly time resolutions ($R^2 = 0.976$, 0.878, and 0.814, respectively). In addition, the WLSTM outperformed the other two widely used algal forecasting approaches - deep neural network (DNN), and autoregressive-integrated-moving-average (ARIMA) model, represented by average 72% and 85% decrease in root-mean-square-error, respectively. Furthermore, the WLSTM was implemented in an experimentally fertilized lake (Lake Tuesday, Michigan) for a multi-step forecasting examination. It satisfactorily forecasted the algal fluctuations involving substantial peak and extreme values (average $R^2 > 0.900$) and presented accurate judgment outcomes to their bloom levels with high accuracy > 95% on average. This work highlighted the utility of deep learning approaches in effective early-warning for algal blooms, and demonstrated an important direction for improving the adaptability of conventional deep learning approaches to the aquatic problems.

1. Introduction

Harmful algal blooms (HABs) have increased worldwide, especially in the regions that are under the intense influence of climate change and human activities (Paerl and Scott, 2010; Reichwaldt and Ghadouani, 2012; Xiao et al., 2019a). Such a HAB event occurring in aquatic habitat may cause critical environmental issues through toxic production and high-biomass accumulation, posing great risks to the ecosystem health and water security (Heisler et al., 2008; Zingone and Oksfeldt Enevoldsen, 2000). The problems arising from HABs have become a globally major concern (Anderson et al., 2002; Hallegraeff, 1993).

Modeling techniques to proactively predict *in-situ* algal dynamics are important for local HAB management. A powerful model can be both beneficial in formulating strategies for the HABs inhibition and in early planning to mitigate negative impacts from a bloom event that is impending or already underway (Coad et al., 2014; Glibert et al., 2010). Nevertheless, considering the complex disturbances of external environments and the non-linear and non-stationary nature of algal dynamics (Li et al., 2015; Zingone and Oksfeldt Enevoldsen, 2000), the forecasting of algal blooms remains a challenging work, particularly for the traditional statistical models based on a priori form lacking self-adaptability (e.g., linear regression, autoregressive moving average models) (Ömer Faruk, 2010; Zhang, 2003).

Machine learning is well-known to be useful in approximating complicated real-world observations (Zhong et al., 2021). Popular algorithms including the neural networks (Deng et al., 2021; Hadjisolomou et al., 2021), support vector regressions (González Vilas et al., 2014; He et al., 2020), random forests (Harris and Graham, 2017; Segura et al., 2017), and gradient boost machines (Xia et al., 2020), have been widely applied in HAB predictions. Recently, given the algorithmic advancements in machine learning, the emergence of deep-learning techniques such as the Long-Short-Term-Memory (LSTM) has received

* Corresponding author.
E-mail address: xi@zju.edu.cn (X. Xiao).

https://doi.org/10.1016/j.watres.2022.118591
Received 15 March 2022; Received in revised form 30 April 2022; Accepted 11 May 2022
Available online 14 May 2022
0043-1354/© 2022 Elsevier Ltd. All rights reserved.

increasing attention in the modeling domain of time series. As an advanced recurrent network, the state-of-the-art LSTM is capable of automatically discovering the underlying temporal dependencies of sequence data through multiple hidden information-processing layers, while avoiding the "gradient vanishing" and "gradient exploding" problems in the recurrent training process (Deng, 2014; Nourani and Behfar, 2021). It was observed that this progress in LSTMs improved the prediction of non-linear algal blooms (Huang et al., 2021; Shin et al., 2019; Zheng et al., 2021). However, pure data-driven approaches are often inadequate to cope with the highly varying periodicity of algal dynamics, i.e., non-stationarity (Wang et al., 2013; Xiao et al., 2017). In practice, an appropriate data pre-processing procedure for the input time series is often needed, which plays an important role in utilizing algorithmic advantages of modeling approaches (Cannas et al., 2006; Nourani et al., 2014).

In recent decades, the highly developed wavelet analysis (WA) has appeared to be practical in addressing the non-stationary difficulty (Nourani et al., 2014). The WA is an advanced time-frequency analysis technique that overcame the drawbacks of classic Fourier analysis, which requires stationary and linear time series data (Labat, 2005). Comparatively, WA is robust to the noisy components and can provide multiple time-scale presentations for the observed signals, which is useful for the interpretation of non-stationary information in a time-series analysis (Daubechies, 1990; Labat, 2005; Li et al., 2015; Nourani and Behfar, 2021). It was shown that the input signal data decomposed by WA could contain more details and could largely benefit the learning of original signals by empirical models (Kisi and Cimen, 2011; Nourani et al., 2014; Xiao et al., 2017). Therefore, in this study, we hypothesized that the hybrid of the deep-learning LSTM algorithm and the time-frequency WA technique (namely, the Wavelet-LSTM; WLSTM) would increase the modeling capacity to capture algal dynamics and thus enhance the forecasting performance of classic LSTMs.

Therefore, the main objective of this work was to explore the algal bloom forecasting by the hybrid WLSTM approach. To achieve the goal: (1) the proposed WLSTM was developed and validated for one-step forecasting of daily algal dynamics in Lake Mendota, Wisconsin; (2) considering the wide temporal variabilities of algal dynamics, the validation was simultaneously conducted at hourly and monthly time resolutions; (3) in addition to the conventional LSTM, the WLSTM was also cross-compared with two other state-of-the-art HAB forecasting approaches at hourly, daily and monthly resolution levels, i.e., deep neural network (DNN) and autoregressive integrated moving average (ARIMA) model; (4) to further test its robustness and serviceability under extreme and frequent bloom conditions, the WLSTM was implemented in an experimentally fertilized lake (Lake Tuesday, Michigan) for a multi-step forecasting examination. The hybrid WLSTM approach could serve as a reliable and cost-effective HAB forecasting tool for water management in the future.

2. Material and methods

2.1. Monitoring data of algal dynamics in the Lake Mendota and the Lake Tuesday

The location of Lake Mendota is near Madison city, southern Wisconsin, USA (Fig. 1). It is a main drainage lake of the Yahara watershed, with a highly developed shoreline of 33.8 km and a surface area of 39.61 km^2. As a eutrophic lake, it has frequently experienced intense phytoplankton blooms over the past decades (Neal, 1987). In 1995, Lake Mendota was added to the North Temperate Lakes-Long Term Ecological Research (NTL-LTER, https://lter.limnology.wisc.edu/) project, and has

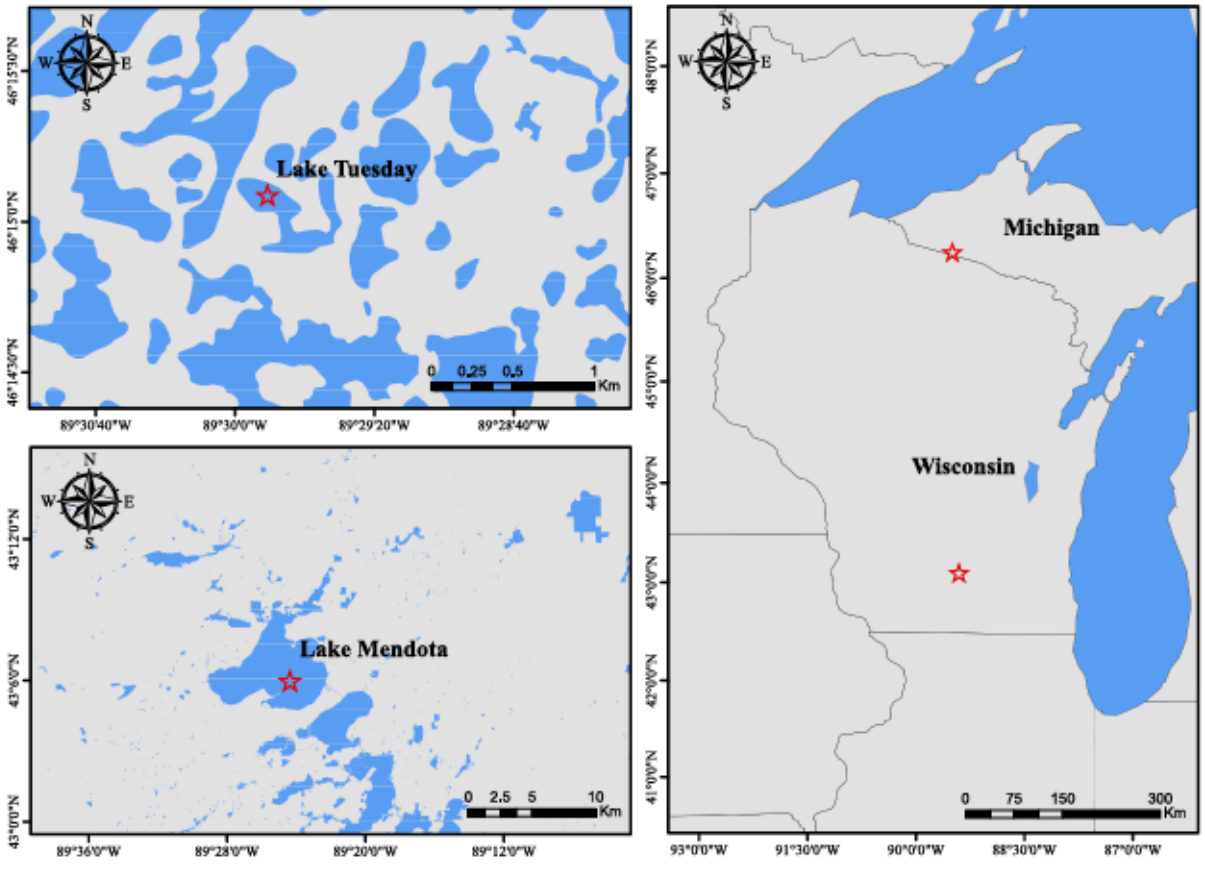

Fig. 1. Locations of monitoring site in the Lake Mendota and the Lake Tuesday.

been continuously monitored at a regular frequency (every 2 weeks during the ice-free season, March to September; every 6 weeks during the ice-covered season) since then. In addition, an instrumented buoy with multiple limnological sensors was deployed on Lake Mendota, which started to provide high-resolution (1 min) monitoring information from 2006, such as the chlorophyll concentration, turbidity, conductivity, pH, and dissolved oxygen. As shown in Table 1, mean monthly cyanobacterial dynamics for a 23-year period (January 1995 to November 2018) were derived from the LTER monitored data (Magnuson et al., 2020a), and were quantified as a natural unit using cell biomass (mg/L). Hourly and daily average values of chlorophyll concentration from 01 Apr. 2020 were calculated from the original high-resolution data (Magnuson et al., 2020b) measured with a multi-parameter analyzer (YSI EXO2, Yellow Springs Instrument Inc., USA), and were reported as relative fluorescence units (RFU) for further development of forecast models.

Lake Tuesday is located in the upper peninsula, Michigan, USA (Fig. 1). It is a relatively small and shallow lake in oligotrophic to mesotrophic conditions before the whole lake manipulation. From 2013 to 2015, a fertilization experiment was conducted there (Pace et al., 2021a). In this experiment, the inorganic nitrogen and phosphorus were gradually added to the lake each year (from mid-May to early September) to cause algal blooms. Meanwhile, the high-frequency chlorophyll fluorometer (Hydrolab DS5X, HYDROLAB Inc., U.S.A) was deployed in this lake to continuously collect chlorophyll-*a* (chl-*a*) concentrations (ug/L) at 5-minute intervals. Similar to the Lake Mendota, daily averaged chl-*a* dynamics were calculated based on the source monitored data (Pace et al., 2021b) and formed the modeling dataset for the Lake Tuesday (Table 1).

Additionally, to determine the outbreak conditions of algal blooms in Lake Tuesday, two chl-*a* thresholds related to the health-based drinking-water supplies were adopted (1 g/L and 12 g/L), as defined by the Alert Level Framework of the World Health Organization (World Health Organization, 2021). The first threshold corresponds to the Alert Level I , representing an early stage of blooms; and the second threshold corresponds to the Alert Level II , representing the heavy stage of bloom situation.

2.2. WLSTM model development

2.2.1. The deep learning algorithm - LSTM network

The LSTM network is a special form of recurrent neural network (RNN), which is a variant of a deep neural network that can learn temporal dependency from sequential data using loop structure. In the LSTM, memory blocks are introduced to replace the hidden neurons for connecting hidden layers (Fig. 2). Each memory block consists of a memory cell (*C*), an input gate (*i*), a forget gate (*f*), and an output gate (*o*). By either updating or removing previously accumulated information to the memory cell through three control gates, the LSTM can well learn the remote dependencies and overcome the gradient exploding and vanishing issues in the backpropagation of conventional RNN (Hochreiter and Schmidhuber, 1997). The learning process and internal structure of the LSTM memory block was presented in Fig. 2a (more algorithmic details and demonstrations were in the supporting information, Text S1), and Fig. 2b showed a basic architecture of LSTM network modeled on the sequential inputs.

To improve the overall performance in deep learning, a standard technique of model ensemble was applied in the LSTM network, where multiple independent models with the same structure were trained and their predictions were combined (Dietterich, 2000). Fig. 2c showed an illustration of the ensemble of LSTM networks.

2.2.2. The time-frequency analysis technique - discrete wavelet transformation

Wavelet transformation (WT) is a useful mathematical tool in signal theory after Fourier transformation, which overcomes the limitations of Fourier transformation in analyzing non-stationary time series (Labat, 2005). By decomposing the main time series into time-frequency space to obtain several sub-series, the WT can effectively extract specific time and frequency features from the original series simultaneously. To achieve this task, the sub-series are generally derived from a template known as the mother wavelet , in which these decomposed wavelets are scaled and translated according to the mother wavelet. Yet, the calculation of scale and translation parameters at every possible position requires substantial computation work for a WT (e.g., continuous wavelet transformation; CWT). Herein, the discrete wavelet transformation (DWT) considerably reduces the computational complexities to apply WT, as its scale and position are usually based on powers of two, called dyadic scales and positions (Cannas et al., 2006). The DWT of series $f\ t$ is often conducted as follows:

(1)

(2)

Where j and k are the integers controlling the decomposition level and translation, respectively; a_0 is the constant scale factor of decomposition (usually is 2), b_0 is the constant position factor of translation (usually is 1); $_{j\ k}\ t$ is the wavelet function; t is the mother wavelet; and $W_f\ j\ k$ are the DWT coefficients. Then a low-frequency approximation sub-series (A_n) and some high-frequency detail sub-series (D_1, D_2, , D_n) can be obtained by repetitively passing the approximation coefficients through the low-pass filter and the high-pass filter at each decomposition level till the preset level is reached.

2.2.3. Hybrid architecture of WLSTM model

The hybrid WLSTM model was developed by combining the wavelet transformation and LSTM network (Fig. 3). The WLSTM modeling mainly has three stages: 1) the DWT of the original series of algal dynamics; 2) the input and forecasting of each sub-series using the ensembled LSTM method individually; 3) and the re-composition of individual output series for the final forecasted results.

2.2.4. Model training and prediction

To appropriately train the deep-learning LSTM modules within the WLSTM structure, our modeling procedure included two phases: (a) calibration and (b) prediction. In the calibration phase, the first 80% of series data were used to construct the networks. To avoid over-fitting problems, the constructing data were randomly divided into 70%

Table 1
Overview of the monitoring datasets for algal dynamics.

Location	Data resolution	Monitoring period	Sample size	Input parameter	Statistical information		
					Mean	S.E.	Range
Lake Mendota	Hourly	2020/04/01 - 2020/7/31	2696	Chlorophyll concentration (RFU)	1.24	0.02	5.72 - 0.01
	Daily	2020/04/01 2020/11/06	211		1.01	0.26	3.83 - 0.03
	Monthly	1995/01 2018/11	219	Cyanobacterial cell biomass (mg/L)	1.56	0.17	14.36 - 0.01
Lake Tuesday	Daily	2013/05/16 2013/09/05	108	Chlorophyll-*a* concentration (ug/L)	14.45	0.58	30.37 7.44
		2014/05/15 2014/09/05	109		16.42	0.45	30.16 9.45
		2015/05/08 2015/08/29	109		15.98	0.52	34.38 8.66

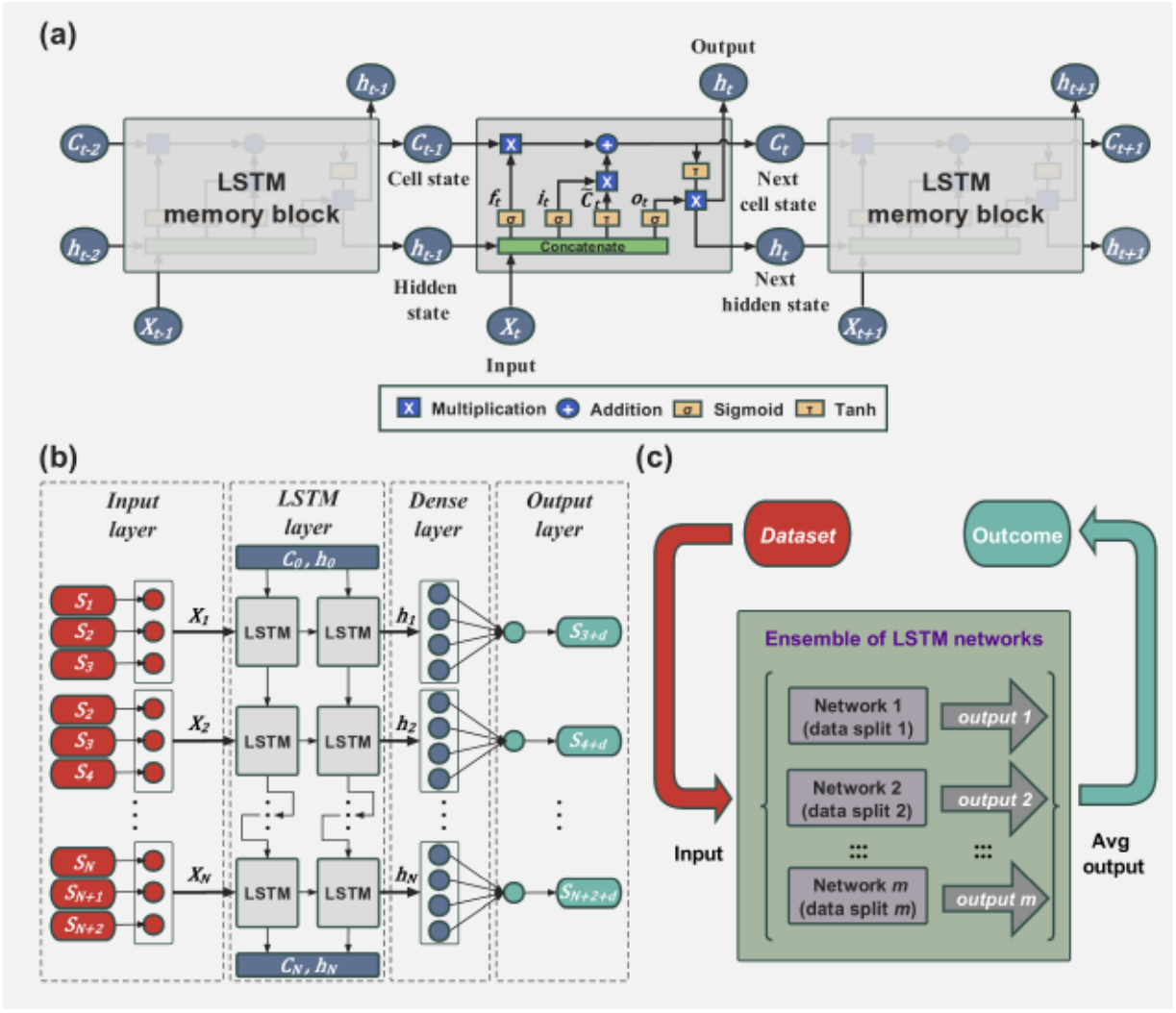

Fig. 2. Sketches of the LSTM models. (a) The learning process and internal structure of LSTM memory block at time *t*. (b) The schematic architecture of the LSTM network with hidden state of 4, LSTM layer of 2, and input dimension of 3. (c) An illustration of the ensemble of LSTM networks. In Fig. 2(b), S_N represents the value of input time series *S* at point *N*, *d* represents the time lag of output time series.

training set and 30% testing set for conducting cross-validation, in which the networks were trained and tested, respectively. After the cross-validation was complete, the model and parameters that performed best on the testing set would be retained. For conducting the ensembling, we repeated this procedure multiple times with each time performed on a different random train-test split of the constructing data and then averaged the outputs (Dietterich, 2000). In the prediction phase, predictions were made on new datasets based on the calibrated models. We implemented both the first 80% of series data (i.e., the entire constructing data) and the remaining 20% of series data (the unused data), to finally evaluate the model performances on in-sample and out-of-sample datasets, respectively. While in the case of Lake Tuesday, 2013 and 2014 data were used for constructing and 2015 data for out-of-sample prediction.

We implemented our LSTM network using the deep-learning framework of PyTorch (Paszke et al., 2017). For computational efficiency, a grid-search scheme based on mini-batch stochastic gradient optimization was used for tuning the network hyperparameters (see Tables S1 & S2). We trained an ensemble of 15 LSTM models according to our prior experiment on the effect of ensemble number on LSTM performance (see Fig S1). In this study, the Daubechies-4 (db4) was applied as a mother wavelet to decompose the original main series into 3 levels due to its wide acceptance and high efficiency (Nourani et al., 2014). The DWT procedures were accomplished in the MATLAB software using the *Wavelet Toolbox*.

2.3. The other time-series forecasting approaches for comparison - DNN and ARIMA

For comparison purposes, the classical deep learning approach – the deep neural network (DNN, i.e., multi-layer feed-forward network with back-propagation algorithm), and the traditional univariate time-series prediction approach – the autoregressive-integrated-moving-average (ARIMA) models, were also developed and tested (detailed information about the two approaches were in Supporting Information, Text S2). To conduct this, the DNN network was trained based on mini-batch stochastic gradient optimization using a grid-search procedure with the *sklearn* library in Python 3.8 software (Pedregosa et al., 2019), and the combinations of model hyperparameters scored the lowest testing loss were lastly selected (see also Tables S1 & S2). We also trained an ensemble of 15 DNN models using the same technique as the LSTM. For the ARIMA approach, models were constructed using the *auto.arima* function of package *forecast* in R environment (Hyndman and Khandakar, 2008), via three iterative steps of model identification, parameter estimation, and model checking.

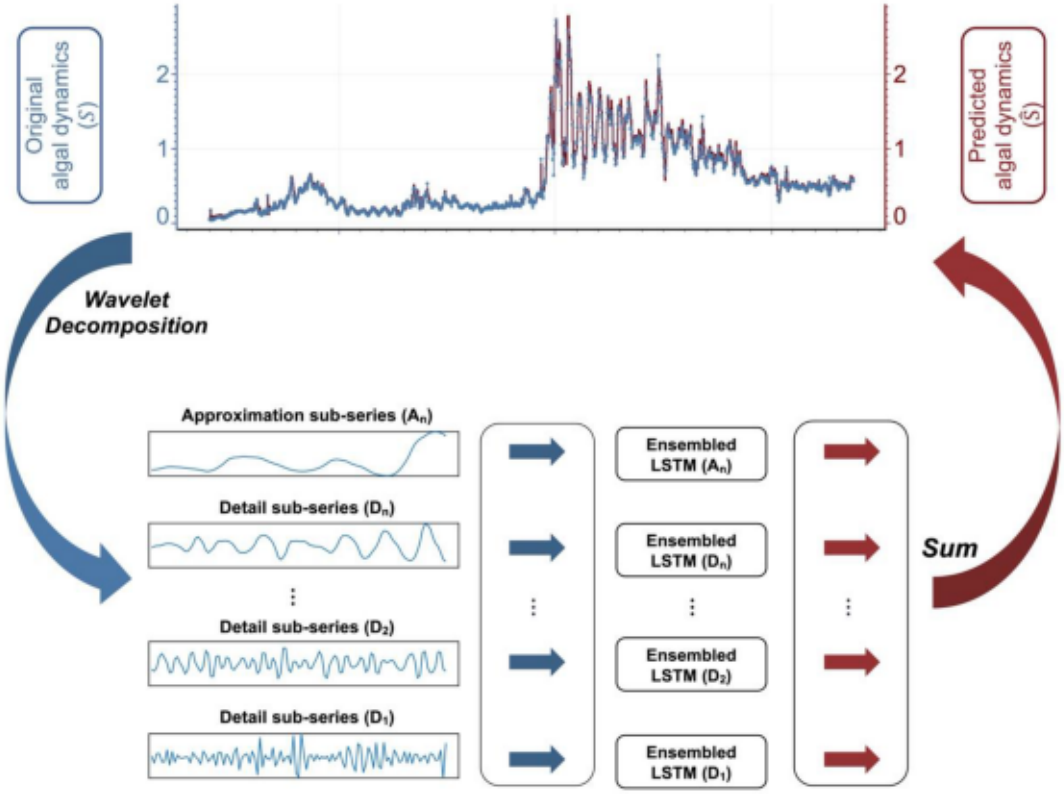

Fig. 3. The forecasting process of the hybrid WLSTM model.

2.4. Data processing

In this study, the inputs of deep-learning models were recomposed from the raw time series (S). We used the first three data points (S_N, S_{N+1}, S_{N+2}) of S to forecast the data points with the time lags of d (S_{N+2+d}) (see Text S3 in the supporting information for more details). The time lag d was set to 1, 2, and 3, respectively, depending on the case studied. For WLSTM modeling, the raw time series (S) was replaced by the decomposed sub-series accordingly (i.e., A_n, D_1, D_2, …, D_n) (Fig. 3).

To facilitate the training of LSTM and DNN models, the input data were pre-normalized. The normalization was denoted by:

$$x_i' = \frac{x_i - \overline{x}}{x_{sd}} \tag{3}$$

Where the x_i' is the normalized value of observed x_i; $\overline{x}$ and x_{sd} are the mean and standard deviation of the observed data, respectively.

2.5. Model assessments

To measure the deviation of predicted values from the observed values, the *correlation coefficient* (R^2), *root-mean-square-error* (RMSE), and *normalized-root-mean-square-error* (NRMSE) were used and calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y} - y_i)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{5}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{6}$$

Where n is the total number of data points; $\hat{y}_i$ and y_i are the i^{th} predicted and observed values; $\overline{y}$ is the mean of y_i; y_{max} and y_{min} are the maxima and minima of y_i.

The predicted values were also used to judge the algal bloom alert levels in Lake Tuesday. Therefore, to evaluate the performance on classification, the *Accuracy* (overall accurate rate), TPR (true positive rate), and TNR (true negative rate) were applied based on the Confusion Matrix, defined as:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{n} \tag{7}$$

$$\mathrm{TPR} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{8}$$

$$\mathrm{TNR} = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \tag{9}$$

Here, n is also the total number of data points.

3. Results

3.1. Forecasting enhancement by the hybrid WLSTM compared to the LSTM

Firstly, the WLSTM approach was developed and verified on the daily algal dynamics in Lake Mendota, Wisconsin. The monitored daily average algal dynamics of Lake Mendota was shown in Fig. 4a (Apr. 1 to Nov. 6, 2020). In detail, the annual algal growth presented a common downward trend, with main peaks concentrated in the first three months (i.e., mid-spring to early summer). In this non-stationary signal mode, although the use of the LSTM model successfully learned algal variations in the calibration phase, the latter prediction showed unsatisfactory forecasting results (Fig. 5b), as indicated by the calculated R^2, RMSE, and NRMSE in Table 2. Furthermore, it can also be seen from Fig. 5b that several sharp changes occurred around the 40th and 80th day causing

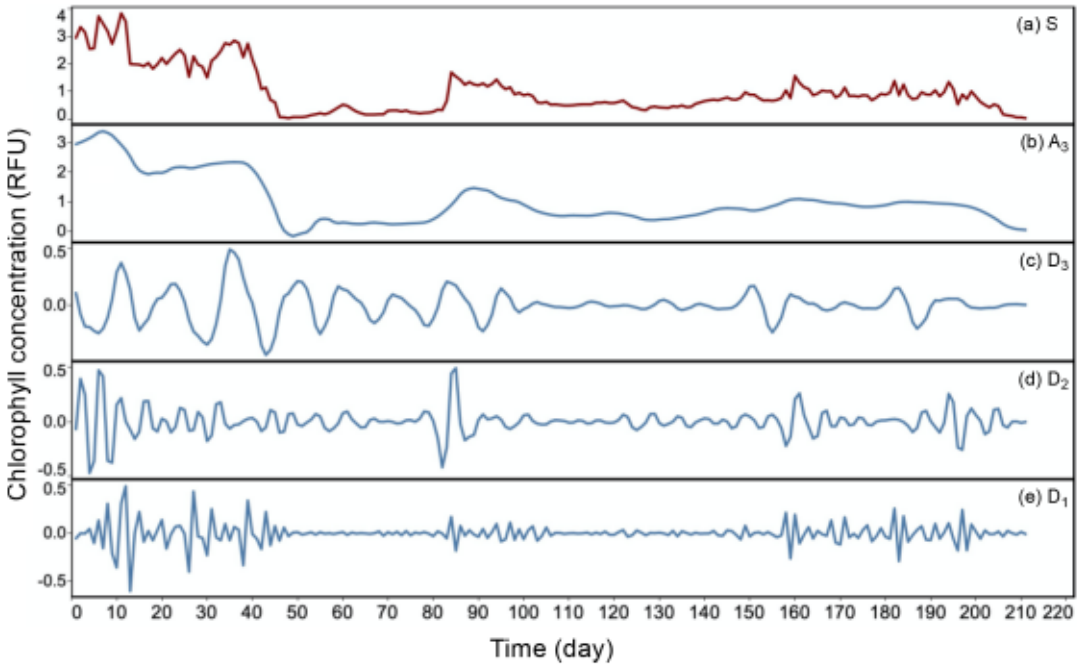

Fig. 4. Wavelet decomposition of the daily average chlorophyll dynamics in the Lake Mendota (S: original main series), using the db4 mother wavelet with approximation coefficient (A_3) and three levels of detailed coefficients (D_1 - D_3).

large simulation errors to the LSTM model, despite it achieving high overall calibration performance.

When coupled with the time-frequency wavelet algorithm, the performance of the WLSTM model significantly improved with an increase of 10% to 57% compared to the original LSTM model (Table 2 & Fig. 5b, Table S3). The wavelet decomposition of daily original algal series (S) usefully produced an approximation coefficient (A_3) and three levels of detailed coefficients (D_1 - D_3) shown in Fig. 4b – 4e. Compared with the main series, the general trend and major peaks of S were identified by the low-frequency sub-series A_3, while its more subtle characteristics were simultaneously captured and presented in the high-frequency sub-series level by level from D_1 to D_3. Moreover, the apparent simulation errors along with the sudden boost or cut of the algal population in the LSTM model were well diminished and smoothed by the hybrid WLSTM method (Fig. 5b).

Noticeably, further validation both proved better forecasting performance of the WLSTM approach at both hourly and monthly time resolutions, as compared to the LSTM approach (Table 2 & Fig. 5). Especially in the long-term monthly validation case (Fig. 5c), the performance of WLSTMs achieved approximately 3 to 6 times higher accuracy than the conventional LSTM in terms of R^2 statistic (0. 858 vs. 0.278 in calibration, and 0.814 vs. 0. 036 in prediction, respectively). Moreover, the WLSTM presented more accurate predictions of the extreme situations that frequently occurred in monthly algal time series (Fig. 5c), with around 60% reduction in the forecasting error (Table S3). In addition, the highest forecasting accuracy of WLSTM was reached at the finest hourly resolution, shown by its biggest R^2 (0.991 and 0.976), smallest RMSE (0.116 and 0.052 RFU), and lowest NRMSE (0.020 and 0.027). All of these evidences indicated that the developed hybrid WLSTM method can reliably forecast the algal dynamics on multiple time scales ranging from hours to months.

3.2. Comparisons of the hybrid WLSTM and other time-series forecasting approaches

The hybrid WLSTM and the other two widely used time-series forecasting approaches (DNN and ARIMA) were also cross-compared based on the algal dynamics in the Lake Mendota (Table 2 & Table S3). The detailed performance of the calibration and prediction via DNN and ARIMA was shown in Fig. S2. For the DNN approach, the best forecasting was achieved at the hourly resolution with highly satisfactory RMSE and NRMSE values, which is consistent with the WLSTM and LSTM approaches. But still, the hourly forecasting performance of DNN dropped by 30% to 90% as compared to WLSTM in terms of RMSE and NRMSE. Interestingly, although the DNN models performed better than LSTM at the monthly resolution (R^2 values of 0.161), it was still largely worse than the WLSTM approach with R^2 value of 0.858, and performed poorly in predicting the extreme values (Fig. S2c). As for the ARIMA approach, all models presented the lowest performance among the three cases studied (Table 2), showing its limitation in coping with non-stationary and complicated forecasting problems of algal dynamics. Furthermore, the recommended ARIMA structure of AR order (p), differencing degree (d), and MA order (q) were the ARIMA (3,1,2), ARIMA (0,1,0), and ARIMA (4,0,2), corresponding to the hourly, daily, and monthly models, respectively (please see Text S4). This indicated that the ARIMA required a longer time series than the other three deep learning network-based approaches to construct a forecasting model. Generally, the WLSTM technique showed huge advantages over the conventional neural network and the statistical ARIMA method.

3.3. Further application of the hybrid WLSTM in an experimental fertilized lake

To test its robustness and serviceability, we further applied the hybrid WLSTM approach in a lake with experimental fertilization (Lake Tuesday) (Fig. 6). In this lake, the excessive nutrient loads apparently triggered the algal over-proliferations, in which the daily chl-*a* concentrations exceeded the early bloom stage threshold all year round (i.e., Alert level I, 1 μg/L). In addition, the large and variational fluctuations of algal dynamics can be observed there (ranging from 7.4 to 34.4 ug/L of chl-*a*), much higher than that of an unmanipulated lake, such as the Lake Mendota with chl-a maximumly ranged from 0.1 to 14.3 mg/L (Fig. 5c & Table 1), leading to an increase of mutation and extreme points within the time series (Fig. 6).

Satisfactory simulations of the algal dynamics were observed in the three WLSTM models with an output time lag of 1, 2, and 3 days,

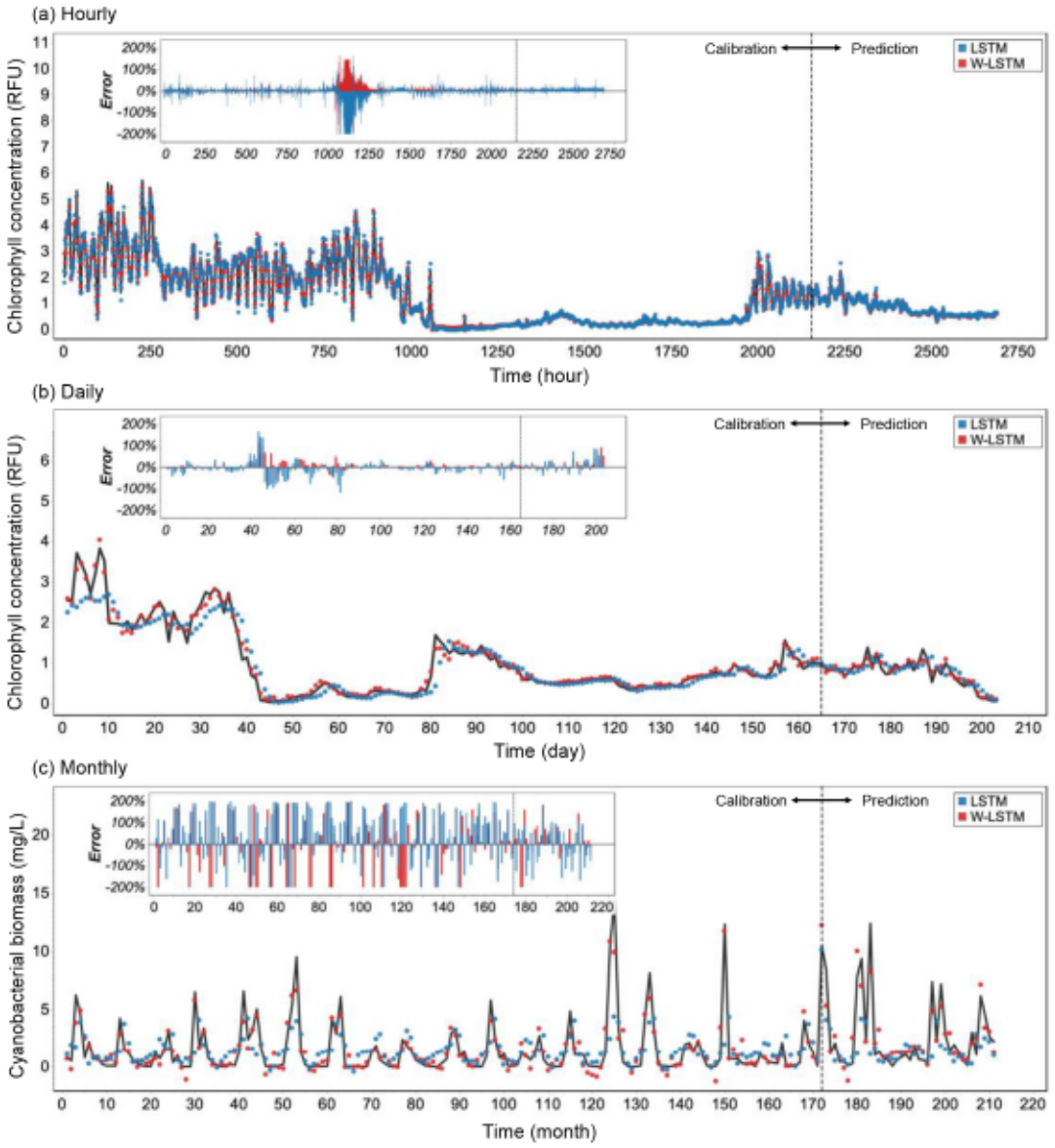

Fig. 5. Observed and forecasted time series of algal biomass by the hybrid WLSTM and conventional LSTM approaches in the Lake Mendota on (a) hourly, (b) daily and (c) monthly resolution for one step ahead. Inner plots represent symmetric relative error (from −200% to 200%).

Table 2
The hourly, daily, and monthly forecasting performance of hybrid WLSTM comparing to LSTM, DNN, and ARIMA in the Lake Mendota.

Resolution	Model	Calibration phase*						Prediction phase*					
		R^2		RMSE		NRMSE		R^2		RMSE		NRMSE	
Hourly	WLSTM	0.991		0.116		0.020		0.976		0.052		0.027	
	LSTM	0.969	(-Δ2.2%)	0.216	(-Δ85.6%)	0.038	(-Δ88.7%)	0.958	(-Δ1.9%)	0.069	(-Δ32.8%)	0.035	(-Δ74.5%)
	DNN	0.970	(-Δ2.1%)	0.214	(-Δ84.4%)	0.037	(-Δ87.4%)	0.958	(-Δ1.9%)	0.069	(-Δ32.8%)	0.035	(-Δ73.0%)
	ARIMA	0.968	(-Δ2.3%)	0.221	(-Δ90.5%)	0.039	(-Δ93.5%)	0.959	(-Δ1.8%)	0.068	(-Δ31.2%)	0.035	(-Δ72.7%)
Daily	WLSTM	0.982		0.115		0.030		0.878		0.114		0.086	
	LSTM	0.904	(-Δ8.0%)	0.269	(-Δ133.9%)	0.071	(-Δ136.4%)	0.615	(-Δ29.9%)	0.202	(-Δ76.8%)	0.153	(-Δ77.8%)
	DNN	0.901	(-Δ8.3%)	0.273	(-Δ137.1%)	0.072	(-Δ139.5%)	0.630	(-Δ28.2%)	0.198	(-Δ73.3%)	0.150	(-Δ74.1%)
	ARIMA	0.911	(-Δ7.3%)	0.269	(-Δ133.6%)	0.071	(-Δ136.4%)	0.550	(-Δ37.4%)	0.224	(-Δ96.8%)	0.166	(-Δ93.3%)
Monthly	WLSTM	0.858		0.855		0.060		0.814		1.382		0.112	
	LSTM	0.278	(-Δ67.6%)	1.924	(-Δ125.0%)	0.134	(-Δ123.3%)	0.036	(-Δ95.6%)	3.057	(-Δ121.2%)	0.264	(-Δ135.7%)
	DNN	0.183	(-Δ78.5%)	2.046	(-Δ139.3%)	0.143	(-Δ137.6%)	0.161	(-Δ80.3%)	2.932	(-Δ112.2%)	0.238	(-Δ112.3%)
	ARIMA	0.264	(-Δ69.2%)	1.928	(-Δ125.5%)	0.134	(-Δ123.3%)	0.006	(-Δ99.3%)	3.130	(-Δ126.5%)	0.253	(-Δ125.9%)

* The values in parentheses represent their relative changes to the WLSTM outcome.

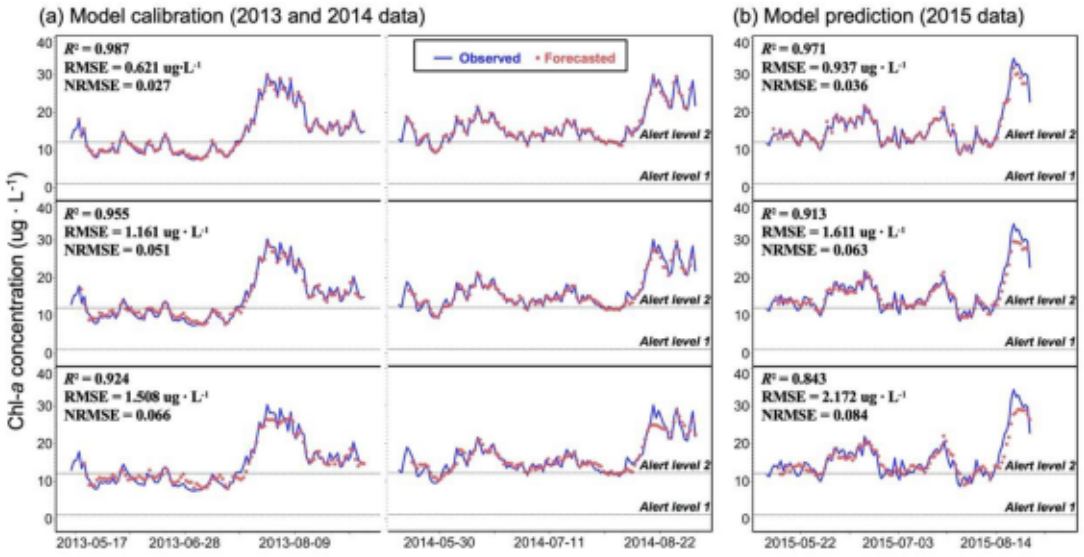

Fig. 6. Performance of the hybrid WLSTM models for forecasting the algal dynamics in the Lake Tuesday, with time lag of 1, 2, and 3 days (from top to bottom, respectively).

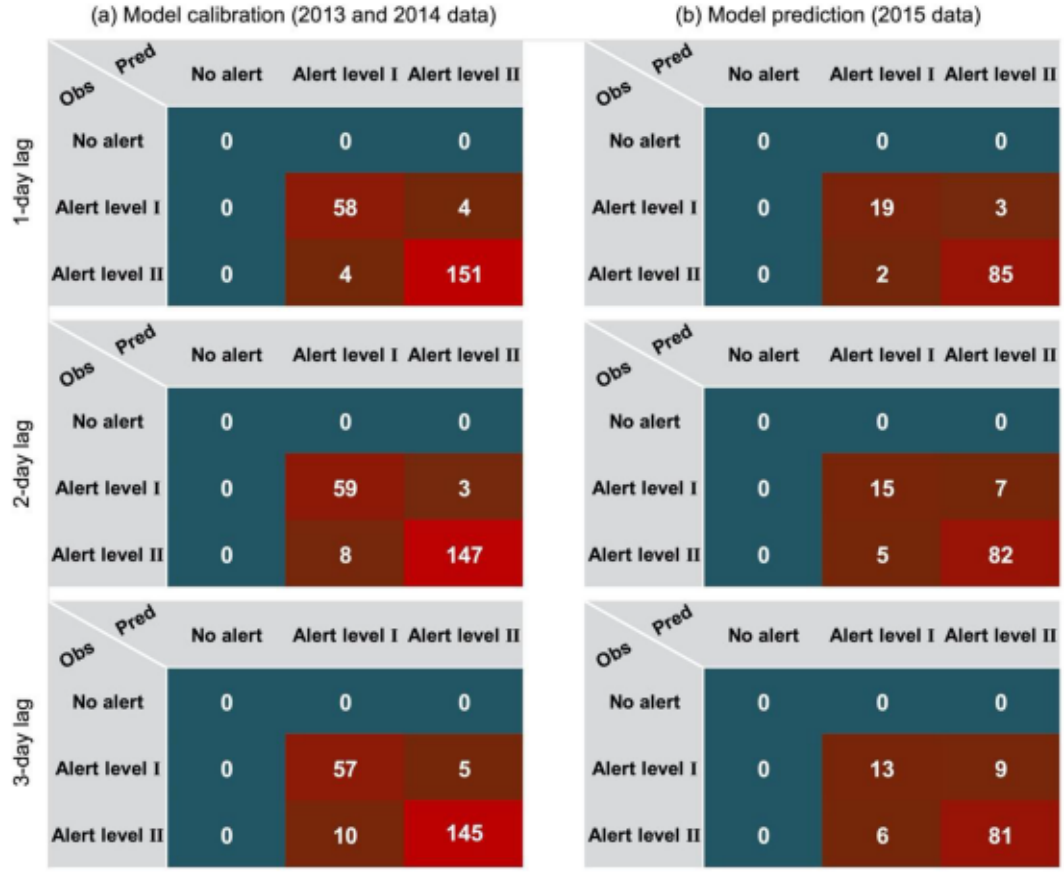

Fig. 7. The exceedance counts of two algal bloom alert thresholds forecasted by the hybrid WLSTM models in the Lake Tuesday, with time lag of 1, 2, and 3 days.

respectively (Figs. 6& 7, Table 3, Table S3). Among three models, the optimal performance was obtained using the forecasting interval of 1 day, as revealed by the highest values of R^2 in calibration (0.987) and prediction (0.971). Nevertheless, it should also be noticed that models for the longer intervals of 2 and 3 days both achieved satisfactory results, which was valuable for the flexible deployment of forecasting work. Furthermore, no clear behavior of model overestimation or underestimation on the peak or extreme values was observed, and all WLSTM models successfully forecasted 217 (100%) exceedances of the level I threshold in the calibration phase and 109 (100%) in the prediction phase (Figs. 6& 7). For the exceedances of level II, an average *Accuracy* rate of 92.5% was accomplished, with TPR > 93.1% ($\overline{TPR}$ = 95.1%) and TNR > 59.3% ($\overline{TNR}$ = 82.8%). Overall, the hybrid WLSTM technique showed a high power in forecasting algal bloom events of the

next multiple time steps.

4. Discussion

Firstly, our results revealed the superiority of artificial-intelligence-based algorithms (LSTM, WLSTM, and DNN) in the HABs forecast, as compared to the traditional linear ARIMA model. Due to various environmental impacts, such as climatic status (Xiao et al., 2019b, 2019a), hydrological condition (Cha et al., 2017; Park et al., 2015), and nutritional level (Beaulieu et al., 2013; Heisler et al., 2008), the dynamics of algal growth were usually shown as a shifty and step-by-step process, characterized by irregular or time-limited time series that was considered to be non-linear and non-stationary (Li et al., 2015; Yu et al., 2020). In our study, the evaluations of ARIMA models (e.g., monthly R^2 0.27, Table 2) indicate that this complicated time-series data is too challenging for the traditional methods, which are often shown to be poor to address HAB forecasting problems (He et al., 2020; Ömer Faruk, 2010; Xiao et al., 2017). Fortunately, the robust performance of deep-learning approaches (LSTM, WLSTM, and DNN) presents promising solutions to this modeling predicament (Table 2). In fact, given the rapid development of such computational intelligence approaches in the ecological field, they have been gradually proven to be well-suited to cope with the complex and real-world aquatic dataset filled with uncertainties (Alizadeh et al., 2018; Kargar et al., 2020; Kouadri et al., 2021; Rousso et al., 2020). Especially, due to the useful characteristics, e.g., the self-learning of sequence dependencies and the efficient structure of gate units, the LSTM approach has been widely suggested to be useful to reliably handle the non-linear HAB dynamics (Cho et al., 2018; Shin et al., 2019; Zheng et al., 2021). However, out of our expectation, the increscent temporal variations and non-stationary oscillations in algal time-series still posed great forecasting difficulties to the advanced LSTM structures. Its prediction R^2 decreased largely from 0.958 (hourly) to 0.615 (daily) and 0.036 (monthly), even lower than the conventional multi-layered network (Table 2).

As we anticipated, the incorporation of the time-frequency WT technique significantly enhanced the LSTM approach (Table 2 & Fig. 5). Compared to the individual LSTM, the forecasting errors of algal dynamics by the hybrid WLSTM were reduced at all three resolutions, with an average decrease of 24%, 44%, and 55% in terms of RMSE, respectively. Specifically, the WT in the WLSTM structure functioned as a powerful pre-processing tool that could remove most of the irregular and noisy components from the raw series, and extract its cyclic signals with expeditious dyadic decompositions (Cannas et al., 2006; Nourani et al., 2014). The extracted sub-series present multi-timescale specialties of the original series periodically or quasi-periodically, facilitating and promoting the algorithmic advantages of LSTM in processing series data. Furthermore, the parallel forecasting process of these sub-series could significantly reduce the error accumulations by the LSTM recursive strategy (Du et al., 2018). This is why in the present study the WLSTM technique can lead to significant increases in the LSTM and outperform the other techniques in algal dynamics prediction (Table 2, Table S3).

More interestingly, the WLSTM model was also observed to be intensively powerful in simulating peak and extreme values with algal dynamics (Figs. 5b & 5c, Figs. S2b & S2c, Table S3). Typically, it is difficult for a data-driven model to accurately predict extreme situations, since this type of model would tend to treat extreme values as outliers before carrying out their normal predictions (Song et al., 2021). However, due to the robust resistance and smoothing capability of WT to the irregular signals, the inclusion of extreme components in decomposed input sub-series can be substantiallydecreased. This not only reduces the probability that WT-coupled models detect the original outliers and thus disrupt the normal prediction, but also again could lead to a great fitting increase in the well-transformed mutations or extreme points (Du et al., 2018; Nourani et al., 2014). As the HAB occurrences are usually accompanied by sharp changes in algal biomass dynamics, the low forecasting error to extreme value is especially useful for the reliable judgment of HAB events. For instance, for the eutrophic water in which high-biomass HABs frequently occur (the Lake Tuesday here), our WLSTM satisfactorily forecasted the highly dynamic process of algal blooms (Fig. 6) and presented accurate judgment outcomes to its bloom levels based on the predicted values (Table 3). Such progress made in HABs models can provide not only more reliable support for the managers but also a better interpretation of this phenomenon in scientific research.

Cross comparison between the time resolution showed that the greatest forecasting improvement of HAB dynamics by WLSTM was in the largest resolution (monthly). Such outcomes, on the one hand, were partially contributed by the unsatisfactory predictions of LSTM models based on a monthly dataset. Compared to the short-term time series in relatively fine time resolutions such as hourly and daily, the long-term monthly series not only vary over a wide scale of time frequencies but could be easily contaminated by intensive cross-periodic oscillations and noises (namely highly non-stationary) (Li et al., 2015). It is evident that the computation intelligence methods still exhibit drawbacks to respond to this circumstance, though they own high capability and flexibility in modeling sequential processes (Cannas et al., 2006; Nourani et al., 2014; Wang et al., 2013). Nevertheless, on the other hand, it is also reported that the underlying periodicity pattern is more dominant in the larger time-scale series than in smaller ones (Nourani et al., 2014). Thus, by decomposing main time series into time-frequency space, the latent modes of periodicity components that existed in such series can be detected and fed to the LSTM to further improve HAB forecasting at a large monthly resolution. This is not alone, in the literature reported by Nourani and Behfar (2021), their WLSTM approach presented analogous modeling results for the runoff-sediment process closed to our time-series patterns of algal dynamics. This again indicates that the WLSTM technique could be useful for developing multi-temporal-scale HAB prediction models.

We also tested the multi-step predictability of hybrid WLSTM models in another fertilized lake (Lake Tuesday). Along with the increase of time lags from 1 day to 3 days, the forecasting accuracy by the WLSTM approach gradually decreased from 0.971 to 0.843 in terms of prediction R^2. The similar phenomenon appeared in many LSTM-based models. For instance, based on the 12-year time series of six water quality parameters, Liang et al. (2020) observed a roughly 50% drop in the performance of LSTM when making predictions for chl-*a* dynamics with the time lag increased from 1 day to 31 days. Zheng et al. (2021) and Shan et al. (2022) also found that the forecasting power of LSTMs reduced in the next multiple time steps (1 h to 12 days), where the RMSE of chl-*a* predictions enlarged by an average of 1.5 to 2 times. This indicated that the autocorrelation structure of algal time series could gradually dispel with the increase of time lag, consequently leading to adverse impacts on the overall modeling accuracy (Shamshirband et al., 2019). In addition, it showed the benefits of automatic capture of long-temporal information via LSTM recurrent chains, which may not successfully work to improve the performance but increase the convergence complexity if the lag time of prediction task is too long (Yang et al., 2019). Therefore, despite a model with multi-step predictability which would be valuable for the practical deployments, selecting an appropriate time lag is still an important concern for the HAB forecasting using WLSTM.

5. Conclusion

In this work, we developed a promising hybrid HAB forecasting approach (WLSTM) combining the wavelet analysis technique (WT) with a deep-learning model (LSTM), which: a) performed better than the state-of-the-art time-series tools including LSTM, DNN, and ARIMA; b) robustly predicted algal dynamics at the hourly, daily, monthly resolutions; c) showed powerful multi-step-ahead predictability and had high overall judgment accuracy to the extreme algal bloom situations over alert thresholds. Therefore, the hybrid approach could be a practical and

Table 3
The accuracy of algal bloom forecasting by the hybrid WLSTM models with time lags varied from 1 to 3 days in the Lake Tuesday.

Alert level	Time lag (day)	Calibration phase*			Prediction phase*		
		TPR	TNR	*Accuracy*	TPR	TNR	*Accuracy*
I	*N* 1	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	N 2	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	N 3	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
II	*N* 1	97.4%	93.5%	96.3%	97.7%	86.4%	95.4%
	N 2	94.8%	95.2%	94.9%	94.3%	68.2%	89.0%
	N 3	93.5%	91.9%	93.1%	93.1%	59.1%	86.3%

* The TPR, TNR, and *Accuracy* represent the true positive rate, true negative rate, and overall accurate rate, respectively.

useful tool for the early warning and management of algal blooms. However, the current research mainly focused on the hybrid use of discrete wavelet transformation and conventional structured LSTMs. In future research, more effective time-frequency analysis techniques and more advanced deep-learning models will be explored to further improve the forecasting accuracy of algal blooms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was financially supported by the National Natural Science Foundation of China (21876148), the Natural Science Foundation of Zhejiang Province/Funds for Distinguished Young Scientists (LR22D06003), the China Postdoctoral Science Foundation (2020M681825), the open founding from the Key Laboratory of Marine Ecological Monitoring and Restoration Technologies of the Ministry of Natural Resources of China (MEMRT202102), Fundamental Research Funds for the Central Universities (226 2022 00119), and Funding for ZJU Tang Scholar (to X. Xiao). We thank PhD student Kokoette Effiong and master student Shitao Huang for their assistances in revision and programing. The authors acknowledge the North Temperate Lakes-Long Term Ecological Research (NTL-LTER, https://lter.limnology.wisc.edu/) project for sharing the monitoring data used in this study. We appreciate those who participated in LTER program by collecting and measuring such high-quality and long-term records for their dedicated work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.watres.2022.118591.

References

Alizadeh, M.J., Kavianpour, M.R., Danesh, M., Adolf, J., Shamshirband, S., Chau, K.W., 2018. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. Eng. Appl. Comput. Fluid Mech. 12, 810 823. https://doi.org/10.1080/19942060.2018.1528480.

Anderson, D.M., Glibert, P.M., Burkholder, J.M., 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25, 704 726. https://doi.org/10.1007/BF02804901.

Beaulieu, M., Pick, F., Gregory-Eaves, I., 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. 58, 1736 1746. https://doi.org/10.4319/lo.2013.58.5.1736.

Cannas, B., Fanni, A., See, L., 2006. Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Phys. Chem. Earth 31, 1164 1171. https://doi.org/10.1016/j.pce.2006.03.020.

Cha, Y., Cho, K.H., Lee, H., Kang, T., Kim, J.H., 2017. The relative importance of water temperature and residence time in predicting cyanobacteria abundance in regulated rivers. Water Res. 124, 11 19. https://doi.org/10.1016/j.watres.2017.07.040.

Cho, H., Choi, U.J., Park, H., 2018. Deep learning application to time-series prediction of daily chlorophyll-a concentration. WIT Trans. Ecol. Environ. 215, 157 163. https://doi.org/10.2495/EID180141.

Coad, P., Cathers, B., Ball, J.E., Kadluczka, R., 2014. Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial neural network. Environ. Model. Softw. 61, 393 409. https://doi.org/10.1016/j.envsoft.2014.07.011.

Daubechies, I., 1990. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf. Theory 36, 961 1005. https://doi.org/10.1109/18.57199.

Deng, L., 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans. Signal Inf. Process. 3 https://doi.org/10.1017/ATSIP.2013.99.

Deng, T., Chau, K.W., Duan, H.F., 2021. Machine learning based marine water quality prediction for coastal hydro-environment management. J. Environ. Manage. 284, 112051 https://doi.org/10.1016/j.jenvman.2021.112051.

Dietterich, T.G., 2000. Ensemble Methods in Machine Learning. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1 15.

Du, Z., Qin, M., Zhang, F., Liu, R., 2018. Multistep-ahead forecasting of chlorophyll a using a wavelet nonlinear autoregressive network. Knowledge-Based Syst. 160, 61 70. https://doi.org/10.1016/j.knosys.2018.06.015.

Glibert, P.M., Allen, J.I., Bouwman, A.F., Brown, C.W., Flynn, K.J., Lewitus, A.J., Madden, C.J., 2010. Modeling of HABs and eutrophication: status, advances, challenges. J. Mar. Syst. 83, 262 275. https://doi.org/10.1016/j.jmarsys.2010.05.004.

González Vilas, L., Spyrakos, E., Torres Palenzuela, J.M., Pazos, Y., 2014. Support Vector Machine-based method for predicting Pseudo-nitzschia spp. blooms in coastal waters (Galician rias, NW Spain). Prog. Oceanogr. 124, 66 77. https://doi.org/10.1016/j.pocean.2014.03.003.

Hadjisolomou, E., Stefanidis, K., Herodotou, H., Michaelides, M., Papatheodorou, G., Papastergiadou, E., 2021. Modelling freshwater eutrophication with limited limnological data using artificial neural networks. Water (Switzerland) 13, 1 15. https://doi.org/10.3390/w13111590.

Hallegraeff, G.M., 1993. A review of harmful algal blooms and their apparent global increase. Phycologia 32, 79 99. https://doi.org/10.2216/i0031-8884-32-2-79.1.

Harris, T.D., Graham, J.L., 2017. Predicting cyanobacterial abundance, microcystin, and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset. Lake Reserv. Manag. 33, 32 48. https://doi.org/10.1080/10402381.2016.1263694.

He, J., Chen, Y., Wu, J., Stow, D.A., Christakos, G., 2020. Space-time chlorophyll-a retrieval in optically complex waters that accounts for remote sensing and modeling uncertainties and improves remote estimation accuracy. Water Res. 171, 115403 https://doi.org/10.1016/j.watres.2019.115403.

Heisler, J., Glibert, P.M., Burkholder, J.M., Anderson, D.M., Cochlan, W., Dennison, W.C., Dortch, Q., Gobler, C.J., Heil, C.A., Humphries, E., Lewitus, A., Magnien, R., Marshall, H.G., Sellner, K., Stockwell, D.A., Stoecker, D.K., Suddleson, M., 2008. Eutrophication and harmful algal blooms: a scientific consensus. Harmful Algae 8, 3 13. https://doi.org/10.1016/J.HAL.2008.08.006.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735 1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water systems. Water Res. 205, 117666 https://doi.org/10.1016/j.watres.2021.117666.

Hyndman, R.J., Khandakar, Y., 2008. Automatic Time Series Forecasting: the forecast Package for R. J. Stat. Softw. 27 https://doi.org/10.18637/jss.v000.i00.

Kargar, K., Samadianfard, S., Parsa, J., Nabipour, N., Shamshirband, S., Mosavi, A., Chau, K.wing, 2020. Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. Eng. Appl. Comput. Fluid Mech. 14, 311 322. https://doi.org/10.1080/19942060.2020.1712260.

Kisi, O., Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J. Hydrol. 399, 132 140. https://doi.org/10.1016/j.jhydrol.2010.12.041.

Kouadri, S., Elbeltagi, A., Islam, A.R.M.T., Kateb, S., 2021. Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast). Appl. Water Sci. 11, 1 20. https://doi.org/10.1007/s13201-021-01528-9.

Labat, D., 2005. Recent advances in wavelet analyses : part 1. A review of concepts. J. Hydrol. 314, 275 288. https://doi.org/10.1016/j.jhydrol.2005.04.003.

Li, W., Qin, B., Zhang, Y., 2015. Multi-temporal scale characteristics of algae biomass and selected environmental parameters based on wavelet analysis in Lake Taihu, China 189 199. https://doi.org/10.1007/s10750-014-2135-7.

Liang, Z., Zou, R., Chen, X., Ren, T., Su, H., Liu, Y., 2020. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. J. Hydrol. 581, 124432 https://doi.org/10.1016/j.jhydrol.2019.124432.

Magnuson, J.J., Carpenter, S.R., Stanley, E.H., 2020a. North Temperate Lakes LTER: phytoplankton - Madison Lakes Area 1995 - current. Environ. Data Initiat. https://doi.org/10.6073/pasta/43d3d401af88cc05c6595962bdb1ab5c.

Magnuson, J.J., Carpenter, S.R., Stanley, E.H., 2020b. North Temperate Lakes LTER: high Frequency Data: meteorological, Dissolved Oxygen, Chlorophyll, Phycocyanin - Lake Mendota Buoy 2006 - current. Environ. Data Initiat. https://doi.org/10.6073/pasta/fc8bd96677405945024ad708003be1fc.

Neal, C., 1987. A eutrophic lake. Chem. Geol. 62, 334 335. https://doi.org/10.1016/0009-2541(87)90098-2.

Nourani, V., Behfar, N., 2021. Multi-station runoff-sediment modeling using seasonal LSTM models. J. Hydrol. 601, 126672 https://doi.org/10.1016/j.jhydrol.2021.126672.

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: a review. J. Hydrol. 514, 358 377. https://doi.org/10.1016/j.jhydrol.2014.03.057.

Ömer Faruk, D., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23, 586 594. https://doi.org/10.1016/j.engappai.2009.09.015.

Pace, M.L., Buelo, C.D., Carpenter, S.R., 2021a. Phytoplankton biomass, dissolved organic matter, and temperature drive respiration in whole lake nutrient additions. Limnol. Oceanogr. 66, 2174 2186. https://doi.org/10.1002/lno.11738.

Pace, M.L., Carpenter, S.R., Buelo, C.D., 2021b. Cascade project at Norther Temperate Lake LTER Daily Respiration Data for Whole Lake Nutrient Additions 2013-2015. Environ. Data Initiat. https://doi.org/10.6073/pasta/9991295661980ff90924f9746b5f42ce.

Paerl, H.W., Scott, J.T., 2010. Throwing fuel on the fire: synergistic effects of excessive nitrogen inputs and global warming on harmful algal blooms. Environ. Sci. Technol. 44, 7756 7758. https://doi.org/10.1021/es102665e.

Park, Y., Pachepsky, Y.A., Cho, K.H., Jeon, D.J., Kim, J.H., 2015. Stressor response modeling using the 2D water quality model and regression trees to predict chlorophyll-a in a reservoir system. J. Hydrol. 529, 805 815. https://doi.org/10.1016/J.JHYDROL.2015.09.002.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., 2017. Automatic differentiation in PyTorch. In: 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA. https://doi.org/10.1145/3434309.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., 2019. Machine learning in Python. J. Mach. Learn. Res. 12, 128 154. https://doi.org/10.4018/978-1-5225-9902-9.ch008.

Reichwaldt, E.S., Ghadouani, A., 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. Water Res. 46, 1372 1393. https://doi.org/10.1016/j.watres.2011.11.052.

Rousso, B.Z., Bertone, E., Stewart, R., Hamilton, D.P., 2020. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. Water Res. 182, 115959 https://doi.org/10.1016/j.watres.2020.115959.

Segura, A.M., Piccini, C., Nogueira, L., Alcántara, I., Calliari, D., Kruk, C., 2017. Increased sampled volume improves Microcystis aeruginosa complex (MAC) colonies detection and prediction using Random Forests. Ecol. Indic. 79, 347 354. https://doi.org/10.1016/j.ecolind.2017.04.047.

Shamshirband, S., Jafari Nodoushan, E., Adolf, J.E., Abdul Manaf, A., Mosavi, A., Chau, K.w., 2019. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. Eng. Appl. Comput. Fluid Mech. 13, 91 101. https://doi.org/10.1080/19942060.2018.1553742.

Shan, K., Ouyang, T., Wang, X., Yang, H., Zhou, B., Wu, Z., Shang, M., 2022. Temporal prediction of algal parameters in Three Gorges Reservoir based on highly time-resolved monitoring and long short-term memory network. J. Hydrol. 605, 127304 https://doi.org/10.1016/j.jhydrol.2021.127304.

Shin, J., Kim, S.M., Son, Y.B., Kim, K., Ryu, J.H., Ryu, J.H., 2019. Early prediction of margalefidinium polykrikoides bloom using a LSTM neural network model in the South Sea of Korea. J. Coast. Res. 90, 236 242. https://doi.org/10.2112/SI90-029.1.

Song, C., Yao, L., Hua, C., Ni, Q., 2021. A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. J. Hydrol. 603, 126879 https://doi.org/10.1016/j.jhydrol.2021.126879.

Wang, F., Wang, X., Chen, B., Zhao, Y., Yang, Z., 2013. Chlorophyll a simulation in a lake ecosystem using a model with wavelet analysis and artificial neural network. Environ. Manage. 51, 1044 1054. https://doi.org/10.1007/s00267-013-0029-5.

World Health Organization, 2021. Chapter 5: exposure to cyanotoxins. In: Chorus, I., Welker, M. (Eds.), Toxic Cyanobacteria in Water, A Guide to Their Public Health Consequences, Monitoring and Management. CRC press, Geneva, p. 859.

Xia, R., Wang, G., Zhang, Y., Yang, P., Yang, Z., Ding, S., Jia, X., Yang, C., Liu, C., Ma, S., Lin, J., Wang, X., Hou, X., Zhang, K., Gao, X., Duan, P., Qian, C., 2020. River algal blooms are well predicted by antecedent environmental conditions. Water Res. 185, 116221 https://doi.org/10.1016/j.watres.2020.116221.

Xiao, X., Agustí, S., Pan, Y., Yu, Y., Li, K., Wu, J., Duarte, C.M., 2019a. Warming amplifies the frequency of harmful algal blooms with eutrophication in Chinese coastal waters. Environ. Sci. Technol. 53, 13031 13041. https://doi.org/10.1021/acs.est.9b03726.

Xiao, X., He, J., Huang, H., Miller, T.R., Christakos, G., Reichwaldt, E.S., Ghadouani, A., Lin, S., Xu, X., Shi, J., 2017. A novel single-parameter approach for forecasting algal blooms. Water Res. 108, 222 231. https://doi.org/10.1016/J.WATRES.2016.10.076.

Xiao, X., He, J., Yu, Y., Cazelles, B., Li, M., Jiang, Q., Xu, C., 2019b. Teleconnection between phytoplankton dynamics in north temperate lakes and global climatic oscillation by time-frequency analysis. Water Res. 154, 267 276. https://doi.org/10.1016/j.watres.2019.01.056.

Yang, D., Chen, K., Yang, M., Zhao, X., 2019. Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features. IET Intell. Transp. Syst. 13, 1475 1482. https://doi.org/10.1049/iet-its.2018.5511.

Yu, Z., Yang, K., Luo, Y., Shang, C., 2020. Spatial-temporal process simulation and prediction of chlorophyll-a concentration in Dianchi Lake based on wavelet analysis and long-short term memory network. J. Hydrol. 582, 124488 https://doi.org/10.1016/j.jhydrol.2019.124488.

Zhang, P.G., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159 175. https://doi.org/10.1016/S0925-2312(01)00702-0.

Zheng, L., Wang, H., Liu, C., Zhang, S., Ding, A., Xie, E., Li, J., Wang, S., 2021. Prediction of harmful algal blooms in large water bodies using the combined EFDC and LSTM models. J. Environ. Manage. 295, 113060 https://doi.org/10.1016/j.jenvman.2021.113060.

Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021. Machine Learning: new Ideas and Tools in Environmental Science and Engineering. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c01339 acs.est.1c01339.

Zingone, A., Oksfeldt Enevoldsen, H., 2000. The diversity of harmful algal blooms: a challenge for science and management. Ocean Coast. Manag. 43, 725 748. https://doi.org/10.1016/S0964-5691(00)00056-9.