Environmentally driven risk assessment for algal bloom occurrence in shallow lakes

Peipei Zhang, Jingqiao Mao∗, Haibin Cai, Wenzin Huang, Shijie Zhu and Tengfei Hu

1 College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing, People’s Republic of China
2 Nanjing Hydraulic Research Institute, Nanjing, People’s Republic of China
* Author to whom any correspondence should be addressed.
E-mail: maojqi@hhu.edu.cn

Keywords: algal blooms, risk assessment, shallow lakes, driving factors, combined influence

Supplementary material for this article is available online

Abstract

An algal bloom is a complex hydro-biological phenomenon driven by multi-attribute environmental processes and thus is still difficult to predict. In this paper, a comprehensive modelling framework for forecasting algal bloom risks in shallow lakes is presented, which is based on long-term field observation and modelling of eutrophic shallow lakes. In the procedure, the major factors and their suitable ranges are investigated, and the individual influence of various driving factors is evaluated quantitatively, using an integrated approach of orthogonal design and regression analysis. By analysing the possible combined effects of the major driving factors and the relationship between algal bloom risk and major bloom-driving factors, a cost-effective environmentally driven risk assessment model is developed to forecast the likelihood of algal bloom occurrence, through a parameter optimization and prediction comparison routine. The risk model has been calibrated and validated against long-term field observations of algal blooms in Taihu Lake, with the prediction accuracy higher than 70%, which only requires readily available meteorological and water quality data. It is noted that for the closed shallow lake, the influence of hydrodynamics can be indirectly reflected by the variation of wind speed; and, total phosphorus, water temperature, photosynthetically active radiation, and average wind speed could be used as major bloom-driving factors in Taihu Lake generally. This study provides a practical framework for the development of algal bloom early warning schemes for shallow lakes and helps to understand the combined function of complex bloom-driving factors.

1. Introduction

The occurrence of eutrophication and algal blooms in inland lakes has been increasing worldwide, which bring a range of hazards including water discoloration, drinking water crisis, hypoxia and fish kills (Guan et al 2022). Algal blooms are a complex hydro-biological phenomenon driven by multi-attribute processes. Multiple factors such as trophic status, physical hydrography and meteorological conditions influence algal growth and abundance. The most commonly observed environmental constraints (or promotions) on algal photosynthesis are nutrients, water temperature (WT) and solar radiation (Gibbons and Bridgeman 2020). And there is a general consensus that the probability of algal bloom occurrence increases in response to increased nutrient loading (Titman 1976, Xu et al 2022). Hydrographic factors (related to the wind for shallow lakes) also play an important role, as most of the algal blooms occur in a weakly flushed water body with relatively long residence times (Mao et al 2014).

Algal blooms can occur and subside over rather short time scales, which are hence notoriously difficult to observe and predict (Janssen et al 2019, Li et al 2021b, Pyo et al 2022). To effectively and proactively reduce hazards caused by algal blooms in freshwater lakes, numerous algal bloom modelling studies have been conducted. Recently, algal blooms are predicted using either process-based (PB) or data-driven (DD)
modelling approaches (Rousso et al 2020). PB models such as PROTECH (Page et al 2018), DYRESM–CAEDYM (Jones et al 2018), and PCLake (Chou et al 2021), are primarily based on some generally accepted framework of eutrophication kinetics, in which the biological and ecological knowledge can be incorporated to solve mass conservation equations numerically. However, their practical application is often hampered by the lack of knowledge of many complex and highly non-linear processes, and the difficulties in the calibration of numerous model parameters (typically over 20–30 parameters). As an alternative approach, DD modelling techniques, such as Regression and Indexes (Mchau et al 2019, Ding et al 2021), Artificial Neural Networks (Yang et al 2018, Park et al 2021) and Bayesian Networks (Mu et al 2021) prospered in algal bloom modelling, combined with remote sensing (Li et al 2022a). DD models are mainly based on data mining algorithms and statistical techniques that analyse and identify patterns amongst monitoring data to create anticipating rules linked to algal bloom dynamics. Nevertheless, since DD models strongly rely on the availability and quality of data, their model performance can be severely constrained if data is not appropriate for the models’ purpose or is not appropriately preprocessed.

For instance, in Taihu Lake, for identification of major driving factors of algal bloom, water quality monitoring and enrichment bioassay experiments (Xu et al 2015); statistical analysis methods (Li et al 2014, Xu et al 2022) and machine learning (Lu et al 2016, Shan et al 2020) have been used extensively to study the relationship between cyanobacterial blooms and the environmental factors in Taihu Lake. However, few studies quantified the individual effect of major driving factors on chlorophyll-a (Chl-a) concentrations (an indicator of algal biomass). For prediction of algal bloom occurrence in Taihu Lake, numerous PB (Huang et al 2012, Li et al 2022b) and DD model (Chen et al 2014, Wang et al 2022) were adopted and proposed, some combined with new technology such as remote sensing (Li et al 2020). However, combined effect of the driving factors usually cannot be fully revealed in data driven models, while the PB models usually are inefficient. A simple and practical model with high accuracy is still needed for effective and proactive algal bloom management in shallow lakes.

In this paper, we develop a comprehensive framework for estimating algal bloom risk in shallow lakes, in which key meteorological, hydrodynamic and biochemical factors, as well as their combined influence are properly accounted for. First, a statistical method is used to identify the major environmental factors driving the algal bloom occurrence. Second, the individual influence of various driving factors is investigated quantitatively. Third, by evaluating the possible combined effect of the major driving factors, a forecast model is developed to forecast the likelihood of algal blooms occurrence. The predicted results are demonstrated, using Taihu Lake as a case study, and the effectiveness and limitation of the proposed modelling framework are finally discussed.

2. Methodology

2.1. Study area

Taihu Lake is the third-largest freshwater lake in China. It is a shallow water body with an average depth of 1.9 m and a surface area of 2338 km$^2$. The lake area is influenced by a humid subtropical climate, with the average annual precipitation about 1000–1400 mm, and the annual average air temperature in 14.9 °C–16.2 °C. The lake can be characterized by several segments, including four major embayments: Zhushan Bay, Meiliang Bay, Gonghu Bay, and Eastern Lake Bay (figure 1). The circulation of this lake is primarily driven by wind, and its hydraulic retention time is around 300 days (if not considering temporary water transfer from the Yangtze River) (Mao et al 2008). The effect of water transfer on water transport processes in the lake is strongly influenced by hydrodynamic conditions induced by wind, and is also influenced by the water transfer routes and flow discharge (Hu et al 2010, Li et al 2013). Taihu Lake was oligotrophic as recently as the 1950s and 1960s, with Chl-a concentrations of about 2 mg m$^{-3}$ and total phosphorus (TP) of about 10 mg m$^{-3}$ (Chen et al 2003). However, since the late 1980s, the water quality of this lake has severely deteriorated and algal blooms have become a frequent nuisance (especially in Meiliang Bay, the most eutrophic region). A series of countermeasures addressing eutrophication and water quality improvement has been implemented since 2007 (cost ~US$14 billion), but nutrient loading and cyanobacterial blooms have not responded as expected to abatement efforts (Qin et al 2019).

2.2. Data collection

The basic geometry and bathymetry data are provided by the Taihu Basin Authority. Monthly samples are taken at 14 monitoring stations across the lake, and the field data from 1995 to 2006, and 2017 are collected to calibrate and validate the forecasting model. For each station, a number of physical and water quality parameters are measured, including Chl-a, WT, pH, chemical oxygen demand (COD$_{\text{Mn}}$), dissolved oxygen (DO), nutrient concentrations of total nitrogen (TN) and TP. Meteorological data of average wind speed (AWS) and photosynthetically active radiation (PAR) are collected at two sites: Taihu Weather Station and Dongshan Weather Station.

2.3. Model development framework

A model development framework is proposed for forecasting algal bloom risks in shallow lakes, using a variety of statistical methods as well as a model
Figure 1. Map of Taihu Lake and distribution of monitoring stations (I-Meiliang Bay; II-Zhushan Bay; III-Gonghu Bay; IV-Western Lake; V-Central Lake; VI-Xuhu Lake; VII-Eastern Lake Bay; VIII-Southern Lake; IX-Wuli Lake).

optimization routine. As shown in figure 2, the proposed framework consists of five steps. This stepwise framework is based on long-term field observation and modelling of the eutrophic shallow lakes. The model input is the monitoring data of meteorological, hydrodynamic and biochemical factors of the lake. The output is the algal bloom risk $R$ calculated from the combine effect model. The intermediate parameters of the model are mainly the weight parameters in the combined effect models in step 4 and are calibrated through an optimization algorithm.

2.4. Combined effect modelling

To quantify the combined effect of factors driving algae growth, the algal bloom risk assessment formula of the additive model is expressed as:

$$R_{\text{add}} = \sum_{i=1}^{G} \omega_{\text{add}}^{i} \text{IE}_{i}$$

where $R_{\text{add}}$ represents algal bloom risk assessed by the additive model; $\text{IE}_{i}$ represents the individual effect of the $i$th major bloom-driving factor; $\omega_{\text{add}}^{i}$ represents the weight of the $i$th major bloom-driving factor ($0 \leq \omega_{\text{add}}^{i} \leq 1$; $\sum_{i=1}^{G} \omega_{\text{add}}^{i} = 1$); $G$ represents the number of the major bloom-driving factors.

The algal bloom risk assessment formula of the multiplicative model is expressed as:

$$R_{\text{mul}} = \prod_{i=1}^{G} \text{IE}_{i}^{\omega_{\text{mul}}^{i}}$$

where $R_{\text{mul}}$ represents bloom risk assessed by the multiplicative model; $\omega_{\text{mul}}^{i}$ represents the weight of the $i$th major bloom-driving factor ($\omega_{\text{mul}}^{i} > 0$); $\text{IE}_{i}$ and $G$ have the same meaning as in equation (1).

We present a composite model to account for the combined effect of the driving factors based on the general mass balance equation for water quality variables (Hydraulics 2006). Theory of the composite model can be found in the Supplementary of this paper. Algal bloom risk assessed by the composite model is represented by $R_{\text{com}}$. The final form of this formula is expressed as:

$$R_{\text{com}} = \omega_{1} \text{IE(PAR)}^{\omega_{\text{PAR}}} \text{IE(WT)}^{\omega_{\text{WT}}} \text{IE(CN)}^{\omega_{\text{CN}}} + \omega_{2} \text{IE(AWS)}^{\omega_{\text{AWS}}}$$

where $\text{IE(PAR)}$ represents the individual effect of PAR; $\text{IE(WT)}$ represents the individual effect of WT; $\text{IE(CN)}$ represents individual effect of the limiting
Figure 2. Forecast model development framework for algal bloom risks in shallow lakes.

nutrient; \(\omega_{\text{PAR}}, \omega_{\text{WT}}, \omega_{\text{CN}}\) are their weights, respectively. Nutrient (TN or TP) with relatively great Pearson correlation coefficient with Chl-a is chosen as \(C_N; \omega_{\text{PAR}}, \omega_{\text{WT}}, \omega_{\text{CN}} > 0\).

To all three risk assessment formulas, the individual effect of each driving factor is normalized by the following equation:

\[
y = 0.1 + 0.8 \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{4}
\]

where \(x\) and \(y\) represents the individual effect of a driving factor before and after the normalization, respectively; \(x_{\text{min}}\) and \(x_{\text{max}}\) represents the minimum and maximum effect of an individual factor before the normalization, respectively.

2.5. Parameter optimization

We develop two objective functions which measure the predictive capability of the three combined effect models. For the optimization model of each risk assessment model, decision variables are weights contained in the corresponding formula and risk thresholds between all bloom grades.

Considering algal bloom divided into \(n\) grades \((1, 2, \ldots, n)\) with their severity in ascending order (non-bloom included and signed as grade 1), \(R_1^i, R_2^i, \ldots, R_n^i\) represent assessed risks of observed blooms of grade \(i\), which are assessed by one of the three risk assessment models, where \(L_i\) represents observed frequency of blooms of grade \(i\). Then the objectives are defined as:
2011b illustrates how the relationships between RT observed bloom of grade 4 and between blooms of grade j upper risk threshold RT between such environmental factors and factors are then selected according to the biochemical mechanisms between such environmental factors and phytoplankton metabolism (Mao et al 2009). Specifically, Chl-a is most strongly related to COD\textsubscript{Mn} and pH, which along with DO can be viewed as consequences of algae growth. Thus, COD\textsubscript{Mn}, pH and DO are excluded to avoid the collinearity among environmental factors. Between the two collected nutrient factors, TN shows a trilling relationship with Chl-a, while TP shows a much stronger relationship. This comparison suggests TN plays a non-limiting role in Taihu Lake. Moderate or strong relationships are established between other bloom-driving factors (i.e. WT, PAR, AWS) and Chl-a. Therefore, TP, WT, PAR, and AWS are identified as major bloom-driving factors.

3.2. Investigation of suitable ranges of major bloom-driving factors
Four levels of the four identified major bloom-driving factors (TP, WT, PAR, and AWS) are examined in 16 experiments under the guidance of the orthogonal array L\textsubscript{4}(4\textsuperscript{3}), as shown in table 1. The median of a series of Chl-a under the designated levels of the four factors according to the observed dataset is summarized in table 2. Four factors affect Chl-a in different ways, with Chl-a ranging from 2.01 to 34.51 µg l\textsuperscript{-1}. The maximum Chl-a occurs in the eighth experimental group with TP, WT, PAR and AWS at level 4, 3, 3 and 2, respectively (i.e. TP is 0.15 ~ 0.80 mg l\textsuperscript{-1}, WT is 18.00 ~ 25.20 °C, PAR is 5.34 ~ 6.86 MJ m\textsuperscript{-2} and AWS is 3.67 ~ 5.34 m s\textsuperscript{-1}). Variance analysis ANOVA (table 3) suggests that the impacts of variations of TP, WT, and AWS on Chl-a are significant (P < 0.01). Moreover, impact of PAR variation is also intense (P < 0.05). The main effect plot (figure 3) is the graphical representation of the results of the range analysis. Depending on figure 3, Chl-a reaches a relatively high value when TP is placed at level 3 and 4 and decreases obviously when TP is at level 1 and 2. Thus, the optimal TP in terms of Chl-a is level 3 and 4, which can be considered as the suitable range of TP. Based on this approach of judgment, suitable ranges of major bloom-driving factors in terms of Chl-a are given in table 4.

3.3. Quantification of individual effects of factors
The use of regression analysis to derive relationships between algal biomass and the four major bloom-driving factors results in four polynomials (table 5). Figure 4 illustrates how the relationships between Chl-a and the four factors vary. According to figure 4, the way in which Chl-a changes with a particular factor is rather similar to that indicated in figure 3. However, when AWS increases from level 1 to level 4, Chl-a declines sharply first then get stable in the latter part based on figure 4, which is not in accordance with the trend presented in figure 3. This may be explained by the lack of representativeness of measured wind speed in the weather stations for the 14 monitoring stations.
Table 1. Levels and factors of the L_{16}(4^5) orthogonal experimental design.

| Factor | TP (mg l^{-1}) | WT (°C) | PAR (MJ m^{-2}) | AWS (m s^{-1}) |
|--------|----------------|---------|-----------------|----------------|
| Level 1 | (0.01, 0.07) | (0.60, 8.80) | (2.24, 3.67) | (1.10, 2.60) |
| Level 2 | (0.07, 0.11) | (8.80, 18.00) | (3.67, 5.34) | (2.60, 3.20) |
| Level 3 | (0.11, 0.15) | (18.00, 25.20) | (5.34, 6.86) | (3.20, 3.50) |
| Level 4 | (0.15, 0.80) | (25.20, 32.00) | (6.86, 10.00) | (3.50, 4.50) |

Note: The significance levels are indicated by *P < 0.05 and **P < 0.01.

Table 2. Orthogonal experiments and median Chl-a.

| Experiment number | Factor | Chl-a (µg l^{-1}) |
|-------------------|--------|------------------|
| 1                 | 1 2 3 3 2 | 4.34 |
| 2                 | 4 1 2 4 2 | 6.64 |
| 3                 | 2 2 4 4 1 | 7.12 |
| 4                 | 3 3 4 1 2 | 34.01 |
| 5                 | 3 1 1 3 1 | 9.63 |
| 6                 | 3 2 2 2 4 | 25.07 |
| 7                 | 2 3 2 3 3 | 13.74 |
| 8                 | 4 3 3 2 1 | 34.51 |
| 9                 | 4 2 1 1 3 | 26.90 |
| 10                | 4 4 4 3 4 | 33.63 |
| 11                | 3 4 3 4 3 | 26.20 |
| 12                | 1 4 2 1 1 | 14.10 |
| 13                | 1 3 1 4 4 | 2.01 |
| 14                | 2 4 1 2 2 | 16.65 |
| 15                | 2 1 3 1 4 | 7.07 |
| 16                | 1 1 4 2 3 | 2.03 |

Table 3. Variance analysis (ANOVA) and significance test of orthogonal experiment results.

| Variance Source | SS    | df | MS    | F     | Significance level |
|-----------------|-------|----|-------|-------|--------------------|
| TP              | 1115.41 | 3 | 371.80 | 144.89 | **                |
| WT              | 648.84 | 3 | 216.28 | 84.28 | **                |
| PAR             | 78.08  | 3 | 26.03  | 10.14 | *                 |
| AWS             | 252.00 | 3 | 84.00  | 32.73 | **                |
| Error           | 7.70   | 3 | 2.57   |       |                    |

Note: The significance levels are indicated by *P < 0.05 and **P < 0.01.

Figure 3. Main effect plot for Chl-a with various levels of TP, WT, PAR and AWS.
### Table 4. Suitable ranges of major bloom-driving factors in terms of Chl-a.

| TP (mg l\(^{-1}\)) | WT (\(^{\circ}\)C) | PAR (MJ m\(^{-2}\)) | AWS (m s\(^{-1}\)) |
|---------------------|---------------------|-----------------------|--------------------|
| [0.11, 0.80]        | [18.00, 32.00]      | [5.34, 10.00]         | [1.1, 3.2)         |

### Table 5. Regression models derived to quantify individual effect of each major factor.

| Factor | Polynomial model for individual effect | \(R^2\) | \(n\) |
|--------|----------------------------------------|---------|-------|
| TP     | Chl-a = 568 TP\(^3\) − 885 TP\(^2\) + 400 TP − 4.20 | 0.895   | 29    |
| WT     | Chl-a = −0.00451 WT\(^3\) + 0.212 WT\(^2\) − 1.45 WT + 16.5 | 0.806   | 19    |
| PAR    | Chl-a = 0.0713 PAR\(^3\) − 1.79 PAR\(^2\) + 17.4 PAR − 21.0 | 0.741   | 20    |
| AWS    | Chl-a = −2.00 AWS\(^3\) + 20.3 AWS\(^2\) − 70.2 AWS + 114 | 0.833   | 15    |

3.4. Calibration and validation of the risk assessment models

Figure 5 shows the Pareto optimal front for APA and MPA calibration for three risk assessment models. The composite model surpasses the other two with respect to both APA and MPA. Built on the property of the Pareto optimal front for the composite model, we set the acceptable MPA as 0.8 then to choose the solution that can achieve higher APA for instance. Consequently, the most desirable solution is chosen with the APA of 0.812 and the MPA of 0.806. Risk threshold between non-bloom and bloom is 0.088. The composite model is finally written as

\[
R_{\text{com}} = 0.892\text{IE}(\text{PAR})^{1.506}\text{IE}(\text{WT})^{0.677}\text{IE}(\text{TP})^{3.694} + 0.108\text{IE}(\text{AWS})^{0.861}.
\]

Model predictions of both bloom and non-bloom occurrence compared with field observations are summarized in table 6. The model has an overall prediction accuracy of (412 + 139)/681 = 0.809. Specifically, the non-bloom prediction accuracy is 0.806. The bloom prediction accuracy is 0.818. And the false alarm rate is as low as 99/511 = 0.194.

Validation of the calibrated composite model is implemented as a follow-up step based on observed data not used in the calibration (table 7). The overall prediction accuracy amounts to 0.736. Prediction accuracy of non-bloom and bloom occurrences are 0.708 and 0.827, respectively. About 92.9% of the 381 predicted non-bloom occurrences are successful. Additionally, a case in April 2017 (18 consecutive days on a monitoring site in the western lake in April 2017) was used to verify the timeliness of the comprehensive model (table 8). In this case, a comparison of simulated and observed bloom and non-bloom occurrences achieved an accuracy of 0.875. Prediction accuracy of non-bloom and bloom occurrences is 0.897 and 0.786, respectively. About 94.5% of the 55
predicted non-bloom occurrences are successful. The results show the timeliness and accuracy of the model prediction.

Figures 6 and 7 show the prediction accuracy of non-bloom and bloom occurrence for different monitoring stations, and temporal variation of Chl-a and bloom risk for different lake regions based on complete datasets during the field observation. It can be seen that the prediction accuracy of bloom occurrence can at least achieve 0.786; two stations (i.e. C1 and C2) are slightly lower for 0.636 and 0.667 respectively. Prediction accuracy of non-bloom occurrence could achieve 0.836, with three stations (i.e. W2, M1, and M3) are a bit lower for the minimum of 0.615. It is noticeable that for stations of W1, W2, and M2, the prediction accuracy of non-bloom occurrence is as low as below 0.62. This may be resulted from the complex hydrodynamics. During the field observation, water diversion project from the Gonghu Bay to Wuli Lake was conducted, increasing the current velocity surrounding W1 and W2. Besides, considering M2 is located near the tributaries of the lake (figure 1), the peripheral region has relatively complicated hydrodynamics.

4. Discussion

4.1. Strengths and limitations

The modelling framework in this study combined the methods of experimental design (the orthogonal design), statistical analysis (correlation and regression), optimization and PB equations (the combined effect model) into a DD process to assess algal bloom risks in shallow lakes. Compared to algal bloom modelling studies using PB models (Huang et al 2012, Li et al 2022b) or DD modelling approaches (Chen et al 2014, Wang et al 2022), it is simple and reveals the algal bloom mechanisms to some degree. To quantify the relationships between the bloom-risking indicator and each driving factor, a factorial design method for experiments is adopted, which helps to select the representative observations and to separate the effects of each individual factor. The composite model based
on the general mass balance equation for water quality variables makes the model equations very simple, along with the timeliness and accuracy of the model prediction. And the framework also provides valuable information on the algal bloom mechanisms when real-time assessing the bloom risks of the lake.

Although the results of the model are promising, there are several limitations in this model framework. Firstly, when applying the modelling framework to different lakes, the essential steps should be redone based on the long-term field observations of environmental factors of the new lake: major bloom-driving factors need to be identified and their suitable ranges reinvestigated correspondingly; and the selected risk assessment model might be different according to the Pareto optimal fronts of the combined effect models. Secondly, the spatial and temporary difference of the algal bloom driving mechanisms in big lakes (Li et al. 2020) could not be addressed at once. Reputation of the modelling work are needed to reveal the characteristics. Thirdly, AWS is used to reflect the influence of hydrodynamics for the shallow lake in this study, while other factors, such as prior wind field, lacustrine topography, hydrophytes and inflow and outflow of the lake are not included. This led to the model not reliable and stable enough in terms of all listed bloom grades and lake regions (see the Model performance section in the supplementary).

4.2. Implications for lake functioning and management

Research on algal blooms of Taihu Lake increased rapidly in recent years, especially after a severe blue-green algae bloom event in 2007. Scholars found that the TP and TN was the most important factor that affected the success of the phytoplankton community in the past 20 years (Deng et al. 2014, Li et al. 2022b). Other physical environment changes such as WT, hydrodynamics, and stratification also directly affect algae growth (Berger et al. 2007, Li et al. 2022a). This is coincident with the major bloom-driving factors identified by the method of this study, and the framework also revealed that the overall effect of TP is more significant than TN and qualified the individual effects of the factors (Paerl et al. 2011, Guo et al. 2017, Li et al. 2022b). The overall prediction accuracy of this model is about 0.809 in calibration and 0.736 in validation, and prediction accuracy of the Chl-a concentration is more than 0.8, satisfactory compared to the existing studies (Liang et al. 2020, Li et al. 2021a).

Results of the case study in Taihu Lake indicates that Eastern Lake, Gonghu Bay, and Southern Lake are lake regions with least bloom occurrences and relatively low bloom risk, followed by Central Lake and Western Lake (figure 7), indicating the northwestern area of Taihu lake is more likely to suffer from algal blooms (Li et al. 2020). For most lake regions the assessed bloom risk complies reasonably with Chl-a, but as observed in figure 7(f), bloom occurrence is often observed with assessed risk below the threshold at WL1. Firstly, 5 of 6 false negatives occur in summer or autumn, when east or southeast wind predominates within the lake area, this might be caused by wind-induced algae drift toward the west or northwest over this period (Kong et al. 2009), could lead to unpredicted bloom occurrences. Secondly, TP of the water inflow in the adjacent tributaries is always higher than that at WL1 during the last three false negatives (no detailed monitoring data for the first three ones), and thus assessed bloom risk at WL1 is not able to represent that in the peripheral region due to TP dilution. This indicates that wind monitoring is important for algal bloom assessment in large shallow...
Figure 7. Temporal variation of Chl-a and bloom risk for eight regions ((a) Meiliang Bay (M4), (b) Central Lake (C2), (c) Wuli Lake (W2), (d) Eastern Lake Bay (E1), (e) Gonghu Bay (G1), (f) Southern Lake (S1), (g) Western Lake (WL1), (h) Western Lake2 (WL2)) of Taihu Lake; shaded region indicates period(s) when assessed risk exceeds the threshold; Chl-a exceeding 30 µg l⁻¹ indicates the bloom condition.

Lakes as Taihu Lake (Chen et al. 2020). Furthermore, seen from the results of major driving factors identification, controlling the TP concentration in Taihu Lake is still the best method for the Chinese government to solve the problem of algal blooms (Xu et al. 2022).
5. Conclusions

The major factors driving the algal blooms (phosphorus, WT, solar radiation, and wind) are statistically identified. The individual effect of various driving factors has been investigated based on an orthogonal experimental design procedure and the regression analysis. By analysing the possible combined effect of the major driving factors, an environmentally driven risk assessment model is developed to forecast the likelihood of algal bloom occurrence with the accuracy higher than 70%. This study provides a comprehensive framework for estimating the algal bloom risk in shallow lakes based on long-term field data.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments

This study received the support of National Key Research and Development Program of China (2021YFC3200301, 2019YFE0109900) and the National Natural Science Foundation of China (52109013, 51909168). The authors would like to thank the Taihu Basin Authority for providing the basic geometry and bathymetry data.

Conflict of interest

We declare that we have no conflict of interest.

References

Berger S A, Diehl S, Stibor H, Trommer G, Ruhestroth M and Wild A 2007 Water temperature and mixing depth affect timing and magnitude of events during spring succession of the plankton Oecologia 150 643–54
Chen H M, Zhu Y Y, Zhang Y, Chen X Q, Wang R C and Zhu W 2020 Cyanobacterial bloom expansion caused by typhoon disturbance in Lake Taihu China Environ. Sci. Pollut. Res. 27 42294–303
Chen Q, Rui H, Li W and Zhang Y 2014 Analysis of algal bloom risk with uncertainties in lakes by integrating self-organizing map and fuzzy information theory Sci. Total Environ. 482 318–24
Chen Y, Fan C, Trautner K and Dokulil M 2003 Changes of nutrients and phytoplankton chlorophyll-a in a large shallow lake, Taihu, China: an 8-year investigation Hydrobiologia 506 273–9
Chen Y, Qin B and Gao X 2001 Prediction of blue-green algae bloom using stepwise multiple regression between algae & related environmental factors in Meiliang Bay Lake Taihu J. Lake Sci. 13 63–71
Chou Q, Nielsen A, Andersen T K, Hu F, Chen W, Zhang X, Cao T, Ni L, Jeppesen E and Troelle D 2021 Assessing impacts of changes in external nutrient loadings on a temperate Chinese drinking water reservoir Front. Environ. Sci. 9 632778
Deng J, Qin B, Paerl H W, Zhang Y, Wu P, Ma J and Chen Y 2014 Effects of nutrients, temperature and their interactions on spring phytoplankton community succession in Lake Taihu, China PLoS One 9 e113960
Ding Y, Zhao J, Peng W, Zhang J, Chen Q, Fu Y and Duan M 2021 Stochastic trophic level index model: a new method for evaluating eutrophication state J. Environ. Manage. 280 111826
Gibbons K J and Bridgeman T B 2020 Effect of temperature on phosphorus flux from anoxic western Lake Erie sediments Water Res. 182 116022
Guan W, Bao M, Lou X, Zhou Z and Yin K 2022 Monitoring, modeling and projection of harmful algal blooms in China Harmful Algae 111 102164
Guo W, Zhu Y, Fu Z, Qin N, Wang H, Liu S, Hu Y, Wu F and Giesy J P 2017 Influences of environmental factors on biomass of phytoplankton in the northern part of Tai Lake, China, from 2000 to 2012 Environ. Monit. Asses. 189 608
Hu L M, Hu W P, Zhai S H and Hu W Y 2010 Effects on water quality following water transfer in Lake Taihu, China Ecol. Eng. 36 471–81
Huang J C, Gao J F and Hormann G 2012 Hydrodynamic-phytoplankton model for short-term forecasts of phytoplankton in Lake Taihu, China Linnmologica 42 7–18
Hydraulics D 2006 Delft3D-WAQ User Manual (The Netherlands: WL Delft Hydraulics)
Janssen A B G et al 2019 How to model algal blooms in any lake on earth Curr. Opin. Environ. Sustain. 36 1–10
Jones H F E, Ozkundakci D, Mcbride C G, Pilditch C A, Allan M G and Hamilton D P 2018 Modelling interactive effects of multiple disturbances on a coastal lake ecosystem J. Environ. Manage. 207 444–55
Kong F, Ma R, Gao J and Wu X 2009 The theory and practice of prevention, forecast and warning on cyanobacteria bloom in Lake Taihu Lake Sci. 3 314–28
Li H, Qin C, He W, Sun F and Du P 2020 Prototyping a numerical model coupled with remote sensing for tracking harmful algal blooms in shallow lakes Glob. Ecol. Conserv. 22 e00938
Li H, Qin C, He W, Sun F and Du P 2021a Improved predictive performance of cyanobacterial blooms using a hybrid statistical and deep-learning method Environ. Res. Lett. 16 124045
Li H, Qin C, He W, Sun F and Du P 2022a Investigating the sub-daily dynamics of cyanobacterial blooms by coupling high-frequency time-series remote sensing with hydro-ecological modelling J. Environ. Manage. 317 10
Li Q, Zhu Y, Chen Q, Li Y, Chen J, Gao Y and Lin P 2021b Spatio-temporal dynamics of water quality and eutrophication in Lake Taihu China Ecosystem 14 e2291
Li S, Liu C, Sun P and Ni T 2022b Response of cyanobacterial bloom risk to nitrogen and phosphorus concentrations in large shallow lakes determined through geographical detector: a case study of Taihu Lake, China Sci. Total Environ. 816 151617
Li Y, Tang C, Wang C, Tian W, Pan B, Hua L, Lau J, Yu Z and Acharya K 2013 Assessing and modeling impacts of different inter-basin water transfer routes on Lake Taihu and the Yangtze River, China Ecol. Eng. 60 399–413
Li Y, Tang C, Yu Z and Acharya K 2014 Correlations between algae and water quality: factors driving eutrophication in Lake Taihu China Int. J. Environ. Technol. 11 169–82
Liang Z, Zou R, Chen X, Ren T, Su H and Liu Y 2020 Simulate the forecast capacity of a complicated water quality model using the long-short-term memory approach J. Hydrol. 581 124152
Liu J, Yang Y S, Gao J and Jiang J 2011a Characteristics of cyanobacteria bloom grading and its temporal and spatial variation in Taihu Lake Resour. Environ. Yangtze Basin 20 156–60 (In Chinese)
Liu J, Yang Y, Jiang J and Gao J 2011b Risk evaluation method of cyanobacteria bloom hazard in Taihu Lake China Environ. Sci. 31 498–503 (In Chinese)
Lu F, Chen Z, Liu W and Shao H 2016 Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin *Ecol. Eng.* **95** 422–9
Mao J, Chen Q and Chen Y 2008 Three-dimensional eutrophication model and application to Taihu Lake China *J. Environ. Sci.* **20** 278–84
Mao J, Jiang D and Dai H 2014 Spatial–temporal hydrodynamic and algal bloom modelling analysis of a reservoir tributary embayment *J. Hydro-Environ. Res.* **9** 296–315
Mao J, Lee J and Choi K W 2009 The extended Kalman filter for forecast of algal bloom dynamics *Water Res.* **43** 4214–24
Mchau G J, Makule E, Machunda R, Gong Y Y and Kimanya M 2019 Phycocyanin as a proxy for algal blooms in surface waters: case study of Ukerewe Island, Tanzania *Water Pract. Technol.* **14** 229–39
Mu M, Li Y, Bi S, Lyu H, Xu J, Lei S, Miao S, Zeng S, Zheng Z and Du C 2021 Prediction of algal bloom occurrence based on the naïve Bayesian model considering satellite image pixel differences *Ecol. Indic.* **124** 107416
Paerl H W, McCarthy M J, Zhu G, Qin B, Li Y and Gardner W S 2011 Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy *Water Res.* **45** 1973–83
Page T, Smith P J, Beven K J, Jones I D, Elliott J A, Maberly S C, Mackay E B, Ville M D and Feuchtmayr H 2018 Adaptive forecasting of phytoplankton communities *Water Res.* **134** 74–85
Park Y, Lee H K, Shin J K, Chon K, Kim S, Cho K H, Kim J H and Baek S S 2021 A machine learning approach for early warning of cyanobacterial bloom outbreaks in a freshwater reservoir *J. Environ. Manage.* **288** 112415
Pyo J, Hong S M, Jang J, Park S, Park J, Noh J H and Cho K H 2022 Drone-borne sensing of major and accessory pigments in algae using deep learning modeling *Gisci Remote Sens.* **59** 310–32
Qin B, Paerl H W, Brookes J D, Liu J, Jeppesen E, Zhu G, Zhang Y, Xu H, Shi K and Deng J 2019 Why Lake Taihu continues to be plagued with cyanobacterial blooms through 10 years (2007–2017) efforts *Sci. Bull.* **64** 354–6
Rousso B Z, Bertone E, Stewart R and Hamilton D P 2020 A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes *Water Res.* **182** 115959
Shan K, Wang X, Yang H, Zhou B, Song L and Shang M 2020 Use statistical machine learning to detect nutrient thresholds in microcystis blooms and microcystin management *Harmful Algae* **94** 101807
Titman D 1976 Ecological competition between algae: experimental confirmation of resource-based competition theory *Science* **192** 463–5
Wang X X, Wang L, Shang M S, Song I R and Shan K 2022 Revealing physiochemical factors and zooplankton influencing Microcystis bloom toxicity in a large-shallow lake using Bayesian machine learning *Toxins* **14** 530
Xu H, Paerl H W, Qin B, Zhu G, Hall N S and Wu Y 2015 Determining critical nutrient thresholds needed to control harmful cyanobacterial blooms in Eutrophic Lake Taihu China *Environ. Sci. Technol.* **49** 1051–9
Xu R, Pang Y, Hu Z and Hu X 2022 The spatiotemporal characteristics of water quality and main controlling factors of algal blooms in Tai Lake, China *Sustainability* **14** 5710
Yang K, Yu Z Y, Luo Y, Yang Y, Zhao L and Zhou X 2018 Spatial and temporal variations in the relationship between lake water surface temperatures and water quality—a case study of Dianchi Lake *Sci. Total Environ.* **624** 859–71