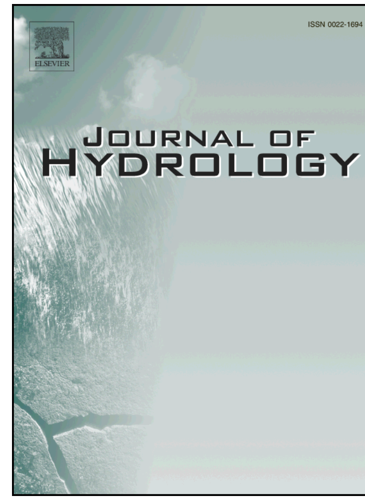
Journal Pre-proofs



Research papers

Establishment of season-specific nutrient thresholds and analyses of the effects of nutrient management in eutrophic lakes through statistical machine learning

Yindong Tong, Xiwen Xu, Shuliang Zhang, Limei Shi, Xiaoyan Zhang, Mengzhu Wang, Miao Qi, Cen Chen, Yingting Wen, Yue Zhao, Wei Zhang, Xuebin Lv

PII:	S0022-1694(19)30814-5
DOI:	https://doi.org/10.1016/j.jhydrol.2019.124079
Reference:	HYDROL 124079
To appear in:	*Journal of Hydrology*
Received Date:	17 May 2019
Revised Date:	20 August 2019
Accepted Date:	26 August 2019

Please cite this article as: Tong, Y., Xu, X., Zhang, S., Shi, L., Zhang, X., Wang, M., Qi, M., Chen, C., Wen, Y., Zhao, Y., Zhang, W., Lv, X., Establishment of season-specific nutrient thresholds and analyses of the effects of nutrient management in eutrophic lakes through statistical machine learning, *Journal of Hydrology* (2019), doi: https://doi.org/10.1016/j.jhydrol.2019.124079

Establishment of season-specific nutrient thresholds and analyses of the effects of nutrient management in eutrophic lakes through statistical machine learning

Authors

Yindong Tong[a], Xiwen Xu[a], Shuliang Zhang[a], Limei Shi[a], Xiaoyan Zhang[a], Mengzhu Wang[a], Miao Qi[a], Cen Chen[b*], Yingting Wen[a], Yue Zhao[c**], Wei Zhang[d], Xuebin Lv[e]

Affiliations

[a] School of Environmental Science and Engineering, Tianjin University, Tianjin, 300072, China

[b] Library, Tianjin University, Tianjin, 300072, China

[c] Chinese Academy for Environmental Planning, Beijing, 100012, China

[d] School of Environment and Natural Resources, Renmin University of China, Beijing, 100872, China

[e] School of Science, Tibet University, Lhasa, 850012, China

Correspondence

*Cen Chen, Email at: yayatulip@163.com;

**Yue Zhao, Email at: zhaoyue@caep.org.cn;

Keywords:

Nutrient thresholds; random forest models; season-specific; nutrient management; lake eutrophication

1. Introduction

Eutrophication and subsequent harmful algal blooms (HABs) have become global water quality problems in recent decades (Conley et al., 2009; Glibert, 2017; Huisman et al., 2018; Paerl et al., 2016a). HABs are usually toxic to humans and other animals, can disrupt aquatic food webs, and result in hypoxia and loss of biodiversity (Peñuelas et al., 2013; Posch et al., 2012; Van de Waal et al., 2014; Zhang et al., 2017). In China, the majority of lakes are rather shallow, and have become or have been becoming eutrophic since the early 2000s (Le et al., 2010). Eutrophication is primarily attributed to intensified anthropogenic nitrogen (N) and phosphorus (P) discharges into freshwater ecosystems (Huisman et al., 2018; Paerl et al., 2016a; Tong et al., 2017a, b; Tong et al., 2018). However, the relative importance of N and P in the control of eutrophication remains a subject that is intensely debated (i.e., the 'P-only' paradigm (Carpenter, 2008; Schindler et al., 2016) versus the 'P + N' paradigm (Lewis et al., 2011; Paerl et al., 2016b). P has traditionally been considered the limiting nutrient for algal growth based on experimental manipulations of lakes in several previous studies (Schindler et al., 2016), and a 'P-only control' strategy has been successful in mitigating eutrophication in some lakes (Lewis et al., 2011). However, increasing numbers of whole-lake experiments have reported that HABs are stimulated by combined P and N enrichment, rather than by enrichment with N or P alone (Paerl et al., 2016b; Paerl et al., 2011). Besides nutrient enrichment, non-manageable environmental factors, such as lake warming, solar radiation, and wind speed, are also believed to be important in explaining the occurrences of HABs (Huisman et al., 2018). Lake warming could promote algal growth, alter algal composition, and increase the concentrations of toxin produced by *Microcystis* spp. and *Planktothrix* spp. (Davis et al., 2009; Paerl et al., 2016a). Reduced water turnover caused by lake warming could be beneficial for buoyant cyanobacteria by allowing them to float upwards (Posch et al., 2012). Despite continuous efforts to explore the relationships between the growth of algae and such environment variables, our current understanding of these relationships is still unclear and inadequate (Robson, 2014; Shimoda and Arhonditsis, 2016).

Several strategies (e.g., nutrient discharge control, increased flushing, chemical treatment, sediment dredging, and aquatic food web manipulation) have been developed to mitigate water eutrophication and control the occurrences of HABs (Huisman et al., 2018; Paerl et al., 2016a). Reduction to the external nutrient discharges into water bodies is believed to ultimately be the most effective control measure, as it addresses the root cause of the problem (Huisman et al., 2018; Tong et al., 2017b; Yu et al., 2019). Before setting a target for nutrient discharge control, establishing accurate nutrient thresholds for defining eutrophication in particular water bodies is crucial (Huo et al., 2018). Due to the huge geographical differences that occur among watersheds (e.g., differences in climate and land use types) and lakes (e.g., with different lake depth and hydrology), ecoregion-specific criteria have been developed to protect water quality by assuming that lakes in the same ecoregion are affected by the same environmental drivers (Cardoso et al., 2007; Poikāne et al., 2010; Richardson et al., 2018). This strategy can make full use of the nutrient monitoring data collected in different lakes, and provides an opportunity for nutrient thresholds to be established for lakes without long-term monitoring data (Huo et al., 2019; Liu et al., 2018). However, recent studies have revealed that nutrient thresholds in lakes could be lake-specific, and thus region-specific nutrient criteria may fail to reflect the natural variations among lakes (Liu et al., 2019; Olson and Hawkins, 2013; Richardson et al., 2018; Rigosi et al., 2014; Taranu et al., 2012). Neglecting such background variations could underprotect water bodies with naturally low nutrient concentrations, but overprotect those with naturally high nutrient concentrations (Olson and Hawkins, 2013). Many studies have also acknowledged the importance of seasonal patterns in environment variables (e.g., water temperature, nutrient concentrations, and solar radiation) to algal growth (Huisman et al., 2018; Paerl et al., 2016a; Posch et al., 2012). However, studies have only rarely addressed the variations in the nutrient thresholds of lakes among different seasons or months.

Establishing nutrient thresholds or criteria for lakes has long been a challenge for water quality managers (Huo et al., 2018). Because nutrients are not toxic to aquatic animals at low concentrations, nutrient criteria cannot be derived based on the does-response relationship used to define threshold levels for toxic pollutants (EPA, 2010). The United States of America (US) was the first country to establish nutrient criteria for waterbodies (EPA, 2000a; Huo et al., 2018). In 2010, three typical methods were recommended by the US Environmental Protection Agency (EPA) to determine the nutrient criteria for particular lakes, which include the reference condition approach, mechanistic models, and stressor-response model (EPA, 2000b; EPA, 2010). In China, official technical guidelines for deriving the nutrient criteria for lakes were not issued until 2017 (Ministry of Ecology and Environment, China, 2019). Among the methods to establish nutrient criteria, the stressor-response model, which describes the most important known relationships between primary productivity and nutrient concentrations, has been the most widely applied in previous studies (Huo et al., 2019; Huo et al., 2018; Liu et al., 2018). The chlorophyll a concentration in a body of water is a water quality index that is closely related to the growth of algae (Liu et al., 2019; Wu et al., 2017; Xu et al., 2015), and is thus usually used as a response variable when establishing nutrient thresholds.

Due to the complicated and sometimes unknown mechanisms involved in algal growth, the capacity of mechanistic water quality modelling to simulate the dynamics of algae that vary over time remains relatively limited (Nelson et al., 2018; Robson, 2014; Shimoda and Arhonditsis, 2016). To study such components of complex biotic community dynamics as the nonlinear and unclear relationships between algal growth and major environmental factors, data-intensive machine learning models (e.g., random forest models, artificial neural networks, support vector machines, etc.) are among the most rigorous tools available (Chou et al., 2018; García Nieto et al., 2019; Liu et al., 2019; Park et al., 2015). These data-intensive machine learning models can achieve even better performance in the simulation of algal growth in lakes than that achieved by traditional mechanistic models (Liu et al., 2019; Nelson et al., 2018). Nelson et al., (2018) applied random forest models to quantify the nature of the relationships between different environmental conditions and five dominant cyanobacterial genera, and estimated the critical nutrient thresholds for different cyanobacterial species. Liu et al., (2019) applied random forest and generalized additive models to assess the predictability of the chlorophyll a concentration in a reservoir and estimate the relative importance of water temperature in driving algal growth. Many similar studies have demonstrated that machine learning methods could effectively simulate algal growth and develop the site-specific nutrient thresholds (Béjaoui et al., 2018; Chou et al., 2018; Park et al., 2015; Shen et al., 2019).

The primary goal of this study was to reveal the potential variations in nutrient thresholds in typical eutrophic lakes among different seasons and assess the potential responses of algal growth to nutrient control methods through the use of machine learning models. Using long-term and unified monitoring datasets (composed of the monthly nutrient monitoring data and meteorological observations collected from 2006 to 2017) from multiple sampling sites in three eutrophic lakes (Lake Taihu, Lake Dianchi, and Lake Chaohu) in China, we applied random forest models to simulate the seasonal algal growth, estimate the critical nutrient thresholds, and assess the potential responses of algal growth to different nutrient control strategies in each lake. The results obtained could offer new insights into how flexible and season-specific nutrient thresholds can be set in eutrophic lakes while accounting for the natural variations in environmental variables, which is crucial for water quality management and reducing the risks of harmful algal blooms in the long-term.

2. Materials and methods

2.1 Lake descriptions

Three typical eutrophic freshwater lakes in China, including Lake Taihu, Lake Chaohu and Lake Dianchi, were selected for examination in this study (Figure S1). These lakes have received much attention from water quality managers in China since the early 2000s because of their serious eutrophication and frequent occurrences of HABs in them (Ministry of Ecology and Environment, China, 2012). Lake Taihu (31.41°N, 120.14°E) is located in the southeastern part of the Yangtze River Basin (Figure S1). It is a large and shallow lake, with an area of 2340 km^2, an average depth of 2.2 m and a water volume of 4.4 billion m^3 (Xu et al., 2015). In Lake Taihu, the most damaging and extensive outbreak of HABs occurred in 2007, which severely affected the water supply of Wuxi City and left over two million people without drinking waters for several weeks (Stone, 2011). Lake Dianchi (25.01°N, 102.66°E) is the largest plateau lake in a traditional phosphate ore mining region of China, with an area of 309 km^2, an average depth of 5.0 m and a water volume of 1.56 billion m^3 (Wu et al., 2017). Lake Chaohu (31.56°N, 117.38°E) is the fifth largest freshwater lake in the lower Yangtze River Basin in China, with an area of 768 km^2, an average depth of 2.7 m and a water volume of 20.7 billion m^3 (Huang et al., 2018).

2.2 Long-term nutrient monitoring and meteorological observational dataset

The data examined in this study consisted of monthly water quality data, including the chlorophyll a (μg/L), total nitrogen (TN, μg/L), total phosphorus (TP, μg/L) and ammonia nitrogen (NH_4^+-N, μg/L) concentrations and Secchi depth (SD, cm) in each lake, and monthly meteorological observations, including daily water temperature (°C), precipitation (mm/day), wind speed (m/s), and sunshine duration (h/day) data, from January 2006 to December 2017. Similarly to previous studies (Chou et al., 2018; Huo et al., 2019; Huo et al., 2018; Liu et al., 2019), we used the chlorophyll a concentration as a proxy for the algal growth in the lakes. Although some other variables could also impact algal growth, to the best of our knowledge, the selected factors examined here included the major potential drivers of algal growth (Huisman et al., 2018; Paerl et al., 2016a), and these variables were also consistently measured with the same standardized methods throughout the study period. In each lake, water quality monitoring was carried out at multiple sampling sites (17 sampling sites in Lake Taihu, 10 in Lake Dianchi and 7 in Lake Chaohu), and the detailed distribution of detailed sampling sites in each lake was shown in Figure S1. Meteorological information was collected from the national meteorological station operated by the

China Meteorological Administration near the lakes (http://data.cma.cn/).

The procedures for collecting water samples and measuring nutrient concentrations were consistent throughout the whole study period and were based on the 'technical specifications requirements for monitoring of surface water and wastewater in China' (HJT 91-2002). Water samples were collected at a depth of about 0.5 m below the water surface. The TN concentration was determined by persulfate digestion, followed by automated colorimetric analysis (N-(1-naphthyl) ethylenediamine dihydrochloride spectrophotometry), with a method detection limit (MDL) of 50 μg/L. The TP concentration was determined by persulfate digestion, followed by automated colorimetric analysis (ammonium molybdate and antimony potassium tartrate under acidic conditions), with an MDL of 10 μg/L. Chlorophyll a concentrations were determined by acetone extraction, followed by separation by centrifugation separation and the determination of sample absorbance, with an MDL of 1 μg/L. All nutrient concentrations lower than the MDL were set to 1/2 of the MDL in subsequent data analyses. The monthly averaged meteorological data were calculated based on the daily observations. In summary, the complete raw dataset comprised a total of 5726 chlorophyll a concentrations (2800 for Lake Taihu, 1430 for Lake Dianchi and 1496 for Lake Chaohu), 5768 TN concentrations (2843 for Lake Taihu, 1429 for Lake Dianchi and 1496 for Lake Chaohu) and 5768 TP concentrations (2843 for Lake Taihu, 1429 for Lake Dianchi and 1496 for Lake Chaohu). A summary of the monitoring data collected during the study period in the three lakes is provided in Table 1.

2.3 Random forest models

We applied random forest models, a typical machine learning method relying on the input of large dataset (Liu et al., 2019; Nelson et al., 2018), to the nutrient monitoring and meteorological observational dataset to characterize the relationships between chlorophyll a concentrations and different environment variables. The input variables in the model include nutrient monitoring data (chlorophyll a, TN, TP, NH_4^+-N and SD) and meteorological data (water temperature, precipitation, wind speed, and sunshine duration) in each lake (Figure 1). The random forest model is a machine learning algorithm that is used to fit a large ensemble of randomly assembled decorrelated classification or regression trees to bootstrapped samples of a response variable, which then averages the outputs of these trees to produce a simulated response (Nelson et al., 2018). The random forest model was previously shown to be good at handling data containing the complicated interactions, and at uncovering the nonlinear and linear relationship structures within such datasets. It has been successfully applied in simulating seasonal algal growth in previous studies (Liu et al., 2019; Nelson et al., 2018). In this study, a random forest model was developed by using the 'randomForest' package in R 3.2.3 and SPSS modeler 18.0 (IBM, USA). Models were performed by the following steps: (1) the historical dataset for each lake during 2006-2016 was portioned into training and testing folds, with 90% of the dataset randomly selected as the training fold that was used to build the random forest model; (2) the random forests were 'grown' through calculations in R or SPSS modeler, and the models' performance was assessed using the testing fold; (3) the models' performance was validated by predicting monthly chlorophyll a concentrations in 2017 and comparing these to observed values; steps 1-3 were repeated for nine times with each new fold representing the 'testing' set in each iteration and (4) the partial dependence of all of the explanatory variables was calculated. The post training model was used to estimate the seasonal nutrient thresholds targeted at different chlorophyll a concentrations (e.g., 10, 20 μg/L, and so on) and assess the potential responses of chlorophyll a concentrations to different nutrient control strategies or water temperatures. When estimating response of chlorophyll a concentrations to reductions of nutrient concentrations in each lake, three scenarios were assumed: 10% reduction in TN (or TP) concentration, 20% reduction in TN (or TP) concentration and 50% reduction in TN (or TP) concentration relative to the monthly monitoring data in 2017. Three scenarios with different water temperatures were assumed: 10% increase, 20% increase and 50% increase relative to the monthly monitoring data in 2017 (Figure 1). The models' performance was quantified using the coefficient of determination (R^2) calculated between the predicted and observed chlorophyll a concentrations. Partial dependence values in the random forest models were also calculated by R 3.2.3 as a measure of each explanatory variable's influence on the response variable given the effects of all the other explanatory variables in the model.

3. Results and discussion

3.1 Summary of long-term nutrient monitoring results

Figure 2 shows the monthly changes in chlorophyll a, TN and TP concentrations from 2006 to 2017 in Lakes Taihu, Dianchi and Chaohu. Among these three lakes, Lake Dianchi had the highest chlorophyll a, TN, and TP concentrations. In 2017, chlorophyll a, TN, and TP concentrations (mean ± standard deviation) in Lake

Dianchi were 86 ± 85, 2226 ± 895 and 137 ± 69 μg/L, respectively, which were all much higher than the corresponding values recorded in Lake Taihu and Lake Chaohu. The TN and TP concentrations in Lake Dianchi were also much higher than the Grade III limit for water quality that is usually used as a standard for clean lakes in China, which is defined as a TN concentration of 1000 μg/L and a TP concentration of 50 μg/L (Ministry of Ecology and Environment, China, 2002). In Lake Taihu and Lake Chaohu, the chlorophyll a concentrations in 2017 were 18 ± 32 and 11 ± 14 μg/L (mean ± standard deviation), respectively, and their TN and TP concentrations approached or were slightly higher than the Grade III limits for lakes in China. During the study period, the TP concentration declined significantly in all three lakes, while the TN concentration was only observed to have declined in Lake Taihu and Lake Dianchi. In response to these changes in TN and TP concentrations, chlorophyll a concentrations gradually declined in Lake Taihu and Lake Chaohu from the year 2006 onward (with a monthly decrease of 0.11 μg/L in Lake Taihu ($P<0.01$, n=144) and of 0.14 μg/L in Lake Chaohu ($P<0.01$, n=144)), while no significant decline in chlorophyll a concentration was observed in Lake Dianchi ($P>0.1$, n=143). Clear seasonal patterns in chlorophyll a and nutrient concentrations were observed in all three lakes (Figure 1). For instance, in Lake Taihu, lower TN concentrations usually occurred in summer, and higher TN concentrations usually occurred in spring, possibly due to changes in internal nutrient cycling (Finlay et al., 2013; Tong et al., 2019; Zhong et al., 2010). In 2017, the TN concentration in July (1118±461 μg/L, mean ± standard deviation) in Lake Taihu was even less than 50% of the TN concentration in this same lake in April (2640±1020 μg/L). Driven by increased water temperatures (Huisman et al., 2018; Paerl et al., 2016a), higher chlorophyll a concentrations usually occurred in summer. Strong spatial variations in nutrient concentrations were observed among different monitoring sites within the same lake (Figure S2). Significant relationships were observed between the TN or TP concentration and the chlorophyll a concentration (*P*<0.01), indicating that nutrients were important drivers of algal growth (Table S1-S3).

3.2 Performance of random forest models

In general, the results of modeling by random forests fit the training data very well. In Lakes Taihu, Dianchi, and Chaohu, the R^2 values calculated for the relationships between the predicted and observed chlorophyll a concentrations in the training datasets were 0.66 ± 0.04 (n=9), 0.79 ± 0.03 (n=9), and 0.73 ± 0.05 (n=9), respectively (Figure S3). In previous studies carried out using linear regression models (LRM) or generalized additive models (GAM), R^2 coefficient values between predicted and observed chlorophyll a concentrations approaching or above 0.2 were believed to indicate effective prediction by the models (Huo et al., 2018; Liu et al., 2018). This indicates that the random forest models used herein could simulate algal growth quite successfully. The results of the comparison of the cross-validated predictions to the testing data are provided in Figure 3A, and these results showed that the random forest models made better predictions of chlorophyll a concentrations in Lake Dianchi (R^2 = 0.48 (0.37 - 0.64), n=9) and Lake Chaohu (R^2 = 0.50 (0.34 - 0.68), n=9) than of those in Lake Taihu (R^2 = 0.26 (0.12 - 0.43), n=9). The models were further validated by predicting the monthly chlorophyll a concentration in each lake in 2017 and comparing it to the observed value. As shown in Figure 3B, the predicted results effectively displayed the correct seasonal variation in chlorophyll a concentrations, and were quite consistent with the observed concentrations in all three lakes (for Lake Taihu, R^2=0.27, P<0.01, n=204; for Lake Dianchi, R^2=0.38, P<0.01, n=120; and for Lake Chaohu, R^2=0.19, P<0.01, n=96; shown in Figure S4).

3.3 Seasonal variations in nutrient thresholds in eutrophic lakes

Partial dependence plots for the random forest models revealed that the relationships between the response variable and the explanatory variable were predominantly nonlinear, and the curves representing these were composed of the average modelled values across the range of observed values of the explanatory variable (Hastie et al., 2009; Nelson et al., 2018). In this study, the steepest curves were associated with the TN concentration, TP concentration, water temperature, and Secchi depth measured in the lakes, while the NH_4^+-N, sunshine duration, precipitation, and wind speed variables were largely invariant (Figure 4). In general, the chlorophyll a concentration had a relatively stronger partial dependence on the TP concentration in these lakes. In particular, these curves indicated that there was a threshold in the relationship between the TP concentration and chlorophyll a concentration, where the partial dependence rose sharply for TP concentrations between 100 and 350 μg/L (Figure 4). In Lake Taihu, this curve plateaued at TP concentrations greater than about 200 μg/L, while in Lake Chaohu and Lake Dianchi, these curves plateaued after about 300 μg/L. Compared with the curves for the TP concentration, the curves between the TN concentration and chlorophyll a concentration were less steep, particularly in Lake Dianchi and Lake Chaohu. With a TN concentration of less than 3000 μg/L, the chlorophyll a concentration remained stable, even as the TN concentrations increased (Figure 4). This result possibly indicates that algae in different lakes could respond differently to the same changes in nutrient concentrations (Olson and Hawkins, 2013; Richardson et al., 2018). As the water temperature rose, the chlorophyll a concentration increased gradually (Figure 4). In Lake Dianchi, a sudden and steep increase in chlorophyll a concentrations occurred at a water temperature of 25 °C, while in Lake Chaohu, a steep increase occurred at a water temperature of 30 °C, suggesting that different algal species could have different sensitivities to increasing water temperatures (Huisman et al., 2018), and that the algae in Lake Dianchi could grow well at mild temperatures (Wang et al., 2019).

By applying the random forest models, we estimated the seasonal variations in the TN and TP thresholds in these lakes that were needed to target different chlorophyll a concentration limits. When estimating thresholds for one variable, we used the monthly monitoring data in these lakes for 2017 as inputs to the models. In general, the estimated TN and TP thresholds varied significantly among different lakes and different months (Figure 5). To limit the chlorophyll a concentration to below 20 μg/L, the estimated monthly TN thresholds in Lakes Taihu, Dianchi, and Chaohu were 2180 ± 479, 2340 ± 295, and 1849 ± 261 μg/L, respectively. The estimated monthly TP thresholds in these lakes were 66 ± 9, 149 ± 28, and 100 ± 22 μg/L, respectively (Figure 5). Significant variations in nutrient thresholds were also observed among different seasons. In Lake Taihu, the TP threshold to limit the chlorophyll a concentration to below 20 μg/L was estimated to be 58 ± 12 μg/L in May, but increased to 82 ± 18 μg/L in September. In Lake Chaohu, the TN threshold was estimated to be 1472 ± 45 μg/L in July, but increased to 2438 ± 200 μg/L in February. The estimated TN and TP thresholds in this study approached the results estimated in a previous study based on bioassay experiments in Lake Taihu (Xu et al., 2015), which were 1260 μg/L for TN and 82 μg/L for TP to limit the chlorophyll a concentration to 20 μg/L in summer. The results of comparing the estimated TN and TP thresholds with the measured results in 2017 showed that the measured TN concentrations approached or were even lower than the estimated thresholds (except for those in the spring in Lake Chaohu). However, the measured TP concentrations were still much higher than the predicted thresholds in summer (Figure 5). For instance, in Lake Taihu, the measured TP concentration (132 ± 51 μg/L) in September was much higher than the estimated threshold value (82 ± 18 μg/L), indicating that the high TP concentrations could possibly be responsible for the high chlorophyll a concentration observed there in summer.

3.4 Responses of chlorophyll a to changes in nutrients and water temperature

Quantifying how algal growth responds to declines in the TN and TP concentrations in lakes is of great importance to setting nutrient control targets for water quality management (Huisman et al., 2018; Paerl et al., 2011; Xu et al., 2015). By applying the random forest models, we estimated the potential declines in chlorophyll a concentrations under scenarios with decreases in TN or TP concentrations of different magnitudes. Figure 6 shows that, the chlorophyll a concentrations in different lakes could have quite different responses to the same decreases in TN and TP concentrations, and greater decreases in chlorophyll a concentrations were observed in scenarios with TP declines than in those with TN declines. With a decline in the TP concentration but no change in the TN concentration, a significant reduction in the chlorophyll a concentration was observed in all three lakes. Larger declines in chlorophyll a concentration were observed with greater decreases in TP concentrations. The largest decrease in chlorophyll a concentration in these scenarios usually occurred in the summer (from July to October) in all three lakes, while only slight changes were observed in spring and winter (Figure 6A). In August, the chlorophyll a concentrations in Lakes Taihu, Dianchi, and Chaohu were predicted to decline from 50 ± 14 to 39 ± 8 μg/L, from 175 ± 120 to 140 ± 100 μg/L, and from 14 ± 18 to 7 ± 1 μg/L, respectively, with a 50% decline in the TP concentration relative to the measured values. However, in spring and winter, the decline in the chlorophyll a concentration in these lakes was usually less than 10% (Figure 6A). Compared with its response to TP concentration, the response of the chlorophyll a concentration to decreases in the TN concentration was not significant (Figure 6B). In all three lakes, the chlorophyll a concentration was only observed to decline as the TN decreased in Lake Chaohu in spring, while only slight changes were observed in the other seasons and lakes. There being different responses of chlorophyll a concentration to nutrient declines in different lakes and seasons revealed the importance of adopting a season-specific nutrient management strategy for controlling the growth of algae. We further estimated the responses of chlorophyll a concentrations to scenarios in which water temperatures increased, and found that the promotion of algal growth by increased water temperature could be quite different in different seasons (Figure 7). In Lake Taihu, a 20% increase in water temperature was estimated to result in a 16% and 23% increase in the chlorophyll a concentration in March and May, respectively, while the corresponding increases in chlorophyll a concentration in July and August were less than 5%, indicating that water temperature might not be a limiting factor for algal growth in this lake in summer (Huisman et al., 2018). In Lake Dianchi, the response of chlorophyll a concentrations to increased temperature was less significant throughout the year (Figure 7), which is consistent with the results presented in Figure 4.

3.5 Implications to future nutrient management in lakes

Although a full mechanistic understanding of the relationships between algal growth and environment variables remains to be attained (Liu et al., 2019; Nelson et al., 2018; Shimoda and Arhonditsis, 2016), the simulated chlorophyll a concentrations produced by the random forest models established in this study were fairly robust (Figure 3, Figures S2 and S3) and demonstrated the importance of establishing lake-specific and season-specific TN or TP thresholds for the control of algal blooms in lakes (Figure 5). Ecoregion-based nutrient criteria provide the possibility of establishing nutrient thresholds for lakes for which long-term nutrient monitoring data are not available (Huo et al., 2019; Huo et al., 2018), and the variations among individual lakes can then represent historical changes in nutrient concentrations (Olson and Hawkins, 2013). Region-based nutrient criteria have been proposed and applied in previously established regional and national nutrient management strategies for lakes (EPA, 2000a; Huo et al., 2018), and they have proven to be effective in water quality protection in some regions (Huo et al., 2018). However, recent studies have revealed that the relative importance of different environmental variables to algal growth could be lake-specific and season-specific, rather than region-specific

(Richardson et al., 2018; Taranu et al., 2012). The natural variations among lakes within the same ecoregion could be so large that the adoption of region-based nutrient criteria could underprotect waterbodies with naturally low nutrient concentrations and overprotect those with naturally high nutrient concentrations (Olson and Hawkins, 2013). For each eutrophic lake selected for use in this study, the estimated nutrient thresholds were quite different, particularly the TP thresholds (Figure 5). The estimated TN nutrient thresholds to limit the chlorophyll a concentration to below 20 μg/L were 2180 ± 479 μg/L for Lake Taihu, 2340 ± 295 μg/L for Lake Dianchi, and 1849 ± 261 μg/L for Lake Chaohu. In Lake Dianchi, the corresponding TP threshold was estimated to be 149 ± 28 μg/L, which was much larger than the estimated values for Lake Taihu (66 ± 9 μg/L, with a range of 56 - 82 μg/L) and Lake Chaohu (100 ± 22 μg/L, with a range of 62 - 126 μg/L). Lake Dianchi is located in a traditional phosphate ore mining area in China, which thus has a naturally high background TP concentration (reaching 100 μg/L as early as 1982) (Ouyang et al., 2015). The previously estimated TN and TP criteria for the region wherein Lake Dianchi was located were estimated to be about 500 μg/L and about 20 μg/L, respectively (Huo et al., 2018). The previously estimated TN and TP criteria for the region wherein Lake Taihu and Lake Chaohu were located were about 360 - 785 μg/L and about 14 - 43 μg/L, respectively (Huo et al., 2018). Because of lack of specific nutrient criteria for many individual lakes, the Grade III limits for TN (1000 μg/L) and TP (50 μg/L) have also been used as the standards for defining clean lakes in China (Ministry of Ecology and Environment, China, 2002; Yu et al., 2019). However, the estimated TN and TP thresholds for the three lakes in this study were much higher than the region-based nutrient criteria and the Grade III limits (Figure 5), which suggests the possibility that these nutrient criteria might have overprotected the water quality of these lakes.

Besides the nutrient enrichment of lakes (Huisman et al., 2018; Paerl et al., 2016a), water temperature is also believed to be a crucial factor determining the algal growth (Huisman et al., 2018; Paerl et al., 2016a). Dimictic lakes usually have a heightened susceptibility to cyanobacterial blooms under stratified eutrophic conditions (Taranu et al., 2012). Lake warming may promote the growth of many bloom-forming species of cyanobacteria and lead to the more stable stratification of the water column and reduced water turnover (Posch et al., 2012). Lake warming could cause changes in algal compositions and further alterations in toxin concentrations (Posch et al., 2012). Different algal species in lakes could have different responses to increasing water temperatures, and toxic *Microcystis* spp. exhibited more significantly elevated growth rates than non-toxic species (Davis et al., 2009). Cyanobacterial species typically reach their maximum growth rates at water temperatures of approximately 30 °C, while chlorophytes and dinoflagellates species usually reach their maximum growth rates at about 25 °C (Paerl et al., 2016a). Due to concerns over the negative impacts of algal blooms caused by climate change, new nutrient management strategies have been proposed, such as setting updated nutrient reduction targets and establishing stricter nutrient criteria for the impacted waterbodies (Huo et al., 2019; Liu et al., 2018). In this study, the analyses of different scenarios carried out using the random forest models showed that the responses of chlorophyll a concentrations to lake warming could differ among different lakes and seasons. Algal growth was not sensitive to lake warming throughout the year in Lake Dianchi. In Lake Taihu, algal production could be promoted significantly by warming in the spring, but not in summer (Figure 7), which indicates the these algal species in this lake may have already reached their maximum growth rate in the summer under present-day conditions (Huisman et al., 2018).

On the other hand, water temperature could also significantly affect the nutrient concentrations within each lake by altering the natural processes occurring therein (e.g., strengthened denitrification, sediment nutrient release, etc.) (Ding et al., 2018; Finlay et al., 2013; Wu et al., 2017), and further can affect the growth of algae (Finlay et al., 2013). Such impacts of water temperature on nutrient concentrations could be particularly important for lakes for which the effective control of anthropogenic nutrient discharges has been established (Wu et al., 2017). In this study, much higher TN concentrations usually occurred in spring and winter, and negative relationships were observed between the water temperature and TN concentration in all three lakes (P<0.01; Figure 8). Higher water temperature is beneficial for the denitrification process, which converts inorganic N species (e.g. NO_3^- and NO_2^-) into N_2 and N_2O, further resulting in decreases in TN concentrations (Yao et al., 2016; Zhong et al., 2010). It was also previously reported that higher water temperature could promote the releases of P from the sediment and increase TP concentrations in water columns, especially in summer (Ding et al., 2018). The strong variations in both TN and TP concentrations within particular lakes could even shift the lakes from following a ‘P-limited pattern’ in spring to an ‘N-limited pattern’ in summer (Xu et al., 2015). This fact indicates that, in addition to promoting algal growth, changes in water temperature could also cause seasonal changes in nutrient levels, which might be good (if they lead to deceasing TN concentrations) or bad (if they lead to increasing TP concentrations) for algal control.

Strong seasonal variations in nutrient concentrations and other environmental drivers of algal growth (e.g., water temperature and solar radiation) require that season-specific, rather than 'one-size fits all', nutrient management strategies are used for eutrophic lakes (Richardson et al., 2018; Taranu et al., 2012). Considering that societies must make decisions based on trade-offs between environmental protection and economic costs, it is necessary to adopt nutrient management strategies based on the monthly nutrient thresholds for bloom-forming cyanobacteria (Yu et al., 2019). Results obtained with the random forest models herein showed that it is more important to control the TP concentration in these lakes than the TN concentration to reduce the chlorophyll a concentrations in therein (Figure 6). The estimated TN and TP thresholds varied significantly among different seasons. In Lake Taihu, the TP criterion to limit the chlorophyll a concentration to below 20 μg/L was estimated to be 58 ± 12 μg/L in May, but increased to 82 ± 18 μg/L in September. In Lake Dianchi, the TP criterion was estimated to be 149 ± 29 μg/L in August, but increased to 202 ± 80 μg/L in April. In Lake Chaohu, the TP criterion was estimated to be 79 ± 29 μg/L in September, but increased to 100 ± 47 μg/L in January (Figure 5). This suggests that it is feasible and necessary to set flexible nutrient criteria in different months, and also that less strict TP criteria might be applied in spring or winter. By comparing the estimated TN and TP criteria with the measured nutrient data in 2017, only the measured TP concentrations in summer were much larger than the estimated nutrient criteria (from July to October), while in other seasons, the measured values approached or were even lower than the estimated thresholds (Figure 5). However, in actual environmental management, the nutrient control target is usually fixed for the same lake throughout the whole year (Huo et al., 2019; Liu et al., 2018), which neglects the natural variations in nutrient concentrations and other environmental variables that occur. This means that many waterbodies are probably overprotected, which increases the economic costs of environmental protection. Specifically, for the studies lakes, the control of TP concentrations in summer could be considered a priority, while the nutrient criteria might be relaxed slightly in other seasons.

4. Conclusion

In this study, we applied random forest models to long-term nutrient monitoring and meteorological observational datasets for Lake Taihu, Lake Dianchi, and Lake Taihu in China. This was done to characterize the relationships between chlorophyll a concentrations and various environmental drivers, establish season-specific nutrient thresholds for each lake, and assess the potential declines in chlorophyll a concentrations that could be achieved through nutrient management. In general, the random forest models performed well at predicting chlorophyll a concentrations, and successfully displayed monthly variations in chlorophyll a concentrations. The estimated TN and TP thresholds were quite variable among different months, and were usually stricter in summer than in winter. To limit chlorophyll a concentrations to remaining below 20 μg/L in August, the estimated TN thresholds in Lakes Taihu, Dianchi, and Chaohu were 2145 ± 683, 2372 ± 918, and 1527 ± 71 μg/L, respectively, and the corresponding TP values were 82 ± 24, 149 ± 22, and 120 ± 22 μg/L. The model results showed that it is more important to control the TP concentration in summer than the TN concentration to reduce the chlorophyll a concentration. The strong seasonal variations in the estimated nutrient thresholds suggest that a 'one-size-fits-all' nutrient control target could overprotect these water bodies and increase the economic costs of eutrophication control. In addition, our results showed that natural changes in water temperature should be considered when establishing such nutrient criteria and establishing a nutrient management strategy.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (41977324, 41630748 and 41671492), Ministry of Science and Technology, China (#2015FY111000) and Tibet University 2018 Central Financial Support Special Funds for Local Colleges and Universities ([2018] No. 54).

References

Béjaoui, B., et al., 2018. Machine learning predictions of trophic status indicators and plankton dynamic in coastal lagoons.Ecol. Indic. 95, 765-774.

Cardoso, A.C., et al., 2007. Phosphorus reference concentrations in European lakes. Hydrobiologia 584(1), 3-12.

Carpenter, S.R., 2008. Phosphorus control is critical to mitigating eutrophication.Proc. Natl. Acad. Sci. U. S. A. 105(32), 11039-11040.

Chou, J.-S., Ho, C.-C., Hoang, H.-S., 2018. Determining quality of water in reservoir using machine learning.Ecol. Inform. 44, 57-75.

Conley, D.J., et al., 2009. Controlling Eutrophication: Nitrogen and Phosphorus. Science 323(5917), 1014-1015.

Davis, T.W., Berry, D.L., Boyer, G.L., Gobler, C.J., 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. Harmful Algae 8(5), 715-725.

Ding, S., et al., 2018. Internal phosphorus loading from sediments causes seasonal nitrogen limitation for harmful algal blooms.Sci. Total Environ. 625, 872-884.

EPA, U., 2000a. Nutrient criteria technical guidance manual: lakes and reservoirs. EPA-822-B-00-001. U.S. Envirpnmental Protection Agency, Office of Water, Washington, DC.

EPA, U., 2000b. Nutrient criteria technical guidance manual, rivers and streems. EPA-822-B-00-002. Office of Water, US Environmental Protection Agence, Washington, DC.

EPA, U., 2010. Using stressor-response relationships to derive numeric nutrient criteria. EPA-820-S-10-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Finlay, J.C., Small, G.E., Sterner, R.W., 2013. Human Influences on Nitrogen Removal in Lakes. Science 342(6155), 247-250.

García Nieto, P.J., García-Gonzalo, E., Alonso Fernández, J.R., Díaz Muñiz, C., 2019. Water eutrophication assessment relied on various machine learning techniques: A case study in the Englishmen Lake (Northern Spain). Ecol. Model.

Glibert, P.M., 2017. Eutrophication, harmful algae and biodiversity — Challenging paradigms in a world of complex nutrient changes.Mar. Pollut. Bull. 124(2), 591-606.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd, ed. Springer Science & Business: New York.

Huang, J., Zhang, Y., Huang, Q., Gao, J., 2018. When and where to reduce nutrient for controlling harmful algal blooms in large eutrophic lake Chaohu, China?Ecol. Indic. 89, 808-817.

Huisman, J., et al., 2018. Cyanobacterial blooms.Nat. Rev. Microbiol. 16(8), 471-483.

Huo, S., et al., 2019. Stricter nutrient criteria are required to mitigate the impact of climate change on harmful cyanobacterial blooms.J. Hydrol. 569, 698-704.

Huo, S., et al., 2018. Development of methods for establishing nutrient criteria in lakes and reservoirs: A review.J. Environ. Sci. 67, 54-66.

Le, C., et al., 2010. Eutrophication of lake waters in China: cost, causes, and control. Environ. Manage. 45(4), 662-668.

Lewis, W.M., Wurtsbaugh, W.A., Paerl, H.W., 2011. Rationale for Control of Anthropogenic Nitrogen and Phosphorus to Reduce Eutrophication of Inland Waters.Environ. Sci. Technol. 45(24), 10300-10305.

Liu, L., et al., 2018. Impacts of climate change and land use on the development of nutrient criteria.J. Hydrol. 563, 533-542.

Liu, X., Feng, J., Wang, Y., 2019. Chlorophyll a predictability and relative importance of factors governing lake phytoplankton at different timescales.Sci. Total Environ. 648, 472-480.

Ministry of Ecology and Environment, China, 2002. Environmental Quality Standards for Surface Water (GB3838-2002).

Ministry of Ecology and Environment, China, 2012. Plan for water pollution control in the key watersheds in China.

Ministry of Ecology and Environment, China, 2019. Technical guideline for deriving nutrient criteria for lakes (HJ 838-2017).

Nelson, N.G., et al., 2018. Revealing Biotic and Abiotic Controls of Harmful Algal Blooms in a Shallow Subtropical Lake through Statistical Machine Learning.Environ. Sci. Technol. 52(6), 3527-3535.

Olson, J.R., Hawkins, C.P., 2013. Developing site-specific nutrient criteria from empirical models. Freshw. Sci. 32(3), 719-740.

Ouyang, Z., Guo, H., Wang, W., Gao, W., 2015. Analysis of Water Quality Change and Impacts from Socio-economic Development in Lake Dianchi from 1982 to 2012 (in Chinses with English abstract). Environmental monitoring in China 31(2), 69-73.

Paerl, H.W., et al., 2016a. Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. Harmful Algae 54, 213-222.

Paerl, H.W., et al., 2016b. It Takes Two to Tango: When and Where Dual Nutrient (N & P) Reductions Are Needed to Protect Lakes and Downstream Ecosystems.Environ. Sci. Technol. 50(20), 10805-10813.

Paerl, H.W., et al., 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(5), 1973-1983.

Park, Y., Cho, K.H., Park, J., Cha, S.M., Kim, J.H., 2015. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea.Sci. Total Environ. 502, 31-41.

Peñuelas, J., et al., 2013. Human-induced nitrogen–phosphorus imbalances alter natural and managed ecosystems across the globe.Nat. Commun. 4, 2934.

Poikāne, S., et al., 2010. Defining Chlorophyll-a Reference Conditions in European Lakes.

Environ. Manage. 45(6), 1286-1298.

Posch, T., Köster, O., Salcher, M.M., Pernthaler, J., 2012. Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming.Nat. Clim. Chang. 2, 809.

Richardson, J., et al., 2018. Effects of multiple stressors on cyanobacteria abundance vary with lake type. Global Change Biol. 24(11), 5044-5055.

Rigosi, A., Carey, C.C., Ibelings, B.W., Brookes, J.D., 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa.Limnol. Oceanogr. 59(1), 99-114.

Robson, B.J., 2014. State of the art in modelling of phosphorus in aquatic systems: Review, criticisms and commentary.Environ. Modell. Softw. 61, 339-359.

Schindler, D.W., Carpenter, S.R., Chapra, S.C., Hecky, R.E., Orihel, D.M., 2016. Reducing Phosphorus to Curb Lake Eutrophication is a Success.Environ. Sci. Technol. 50(17), 8923-8929.

Shen, J., Qin, Q., Wang, Y., Sisson, M., 2019. A data-driven modeling approach for simulating algal blooms in the tidal freshwater of James River in response to riverine nutrient loading.Ecol. Model. 398, 44-54.

Shimoda, Y., Arhonditsis, G.B., 2016. Phytoplankton functional type modelling: Running before we can walk? A critical evaluation of the current state of knowledge.Ecol. Model. 320, 29-43.

Stone, R., 2011. China Aims to Turn Tide Against Toxic Lake Pollution. Science 333(6047), 1210-1211.

Taranu, Z.E., Zurawell, R.W., Pick, F., Gregory-Eaves, I., 2012. Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. Global Change Biol. 18(12), 3477-3490.

Tong, Y., et al., 2017a. Estimation of nutrient discharge from the Yangtze River to the East China Sea and the identification of nutrient sources.J. Hazard. Mater. 321, 728-736.

Tong, Y., et al., 2017b. Decline in Chinese lake phosphorus concentration accompanied by shift in sources since 2006.Nat. Geosci. 10, 507-511.

Tong, Y., et al., 2019. Impacts of water residence time on nitrogen budget of lakes and reservoirs.Sci. Total Environ. 646, 75-83.

Tong, Y., et al., 2018. Human activities altered water N:P ratios in the populated regions of China. Chemosphere 210, 1070-1081.

Van de Waal, D.B., Smith, V.H., Declerck, S.A.J., Stam, E.C.M., Elser, J.J., 2014. Stoichiometric regulation of phytoplankton toxins.Ecol. Lett. 17(6), 736-742.

Wang, J.-H., et al., 2019. Meteorological factors and water quality changes of Plateau Lake Dianchi in China (1990–2015) and their joint influences on cyanobacterial blooms.Sci. Total Environ. 665, 406-418.

Wu, Z., Liu, Y., Liang, Z., Wu, S., Guo, H., 2017. Internal cycling, not external loading, decides the nutrient limitation in eutrophic lake: A dynamic model with temporal Bayesian hierarchical inference.Water Res. 116, 231-240.

Xu, H., et al., 2015. Determining Critical Nutrient Thresholds Needed to Control Harmful Cyanobacterial Blooms in Eutrophic Lake Taihu, China.Environ. Sci. Technol. 49(2), 1051-1059.

Yao, L., Jiang, X., Chen, C., Liu, G., Liu, W., 2016. Within-lake variability and environmental controls of sediment denitrification and associated N2O production in a shallow eutrophic lake.Ecol. Eng. 97, 251-257.

Yu, C., et al., 2019. Managing nitrogen to restore water quality in China. Nature 567(7749), 516-520.

Zhang, Y., et al., 2017. Global loss of aquatic vegetation in lakes.Earth-Sci. Rev. 173, 259-265.

Zhong, J., et al., 2010. Seasonal variation of potential denitrification rates of surface sediment from Meiliang Bay, Taihu Lake, China.J. Environ. Sci. 22(7), 961-967.

Figure captions

Figure 1. Flowchart of model development and application in predicting nutrient thresholds and response to different nutrient reductions;

Figure 2. Monthly changes of TN, TP and chlorophyll a concentration in Lake Taihu, Dianchi and Chaohu between 2006 and 2017;

Figure 3. Overall cross-validated predictions by the random forest model in Taihu, Chaohu and Dianchi. A. Comparison between predicted chlorophyll a concentration and measured value in the testing data set; B. Prediction of monthly chlorophyll a concentration in three lakes in 2017;

Figure 4. Partial dependence plots of chlorophyll a concentration to different explanatory variables in the random forest model

Figure 5. Estimated monthly nutrient thresholds for (A) TP concentration and (B) TN concentration in Taihu, Dianchi and Chaohu;

Figure 6. Response of chlorophyll a concentration to decrease of TP (A) and TN (B) concentrations in the lakes;

Figure 7. Response of chlorophyll a concentration to the increases of water temperatures in three lakes;

Figure 8. Correlation between water temperature and TN concentrations in Taihu, Dianchi and Chaohu;

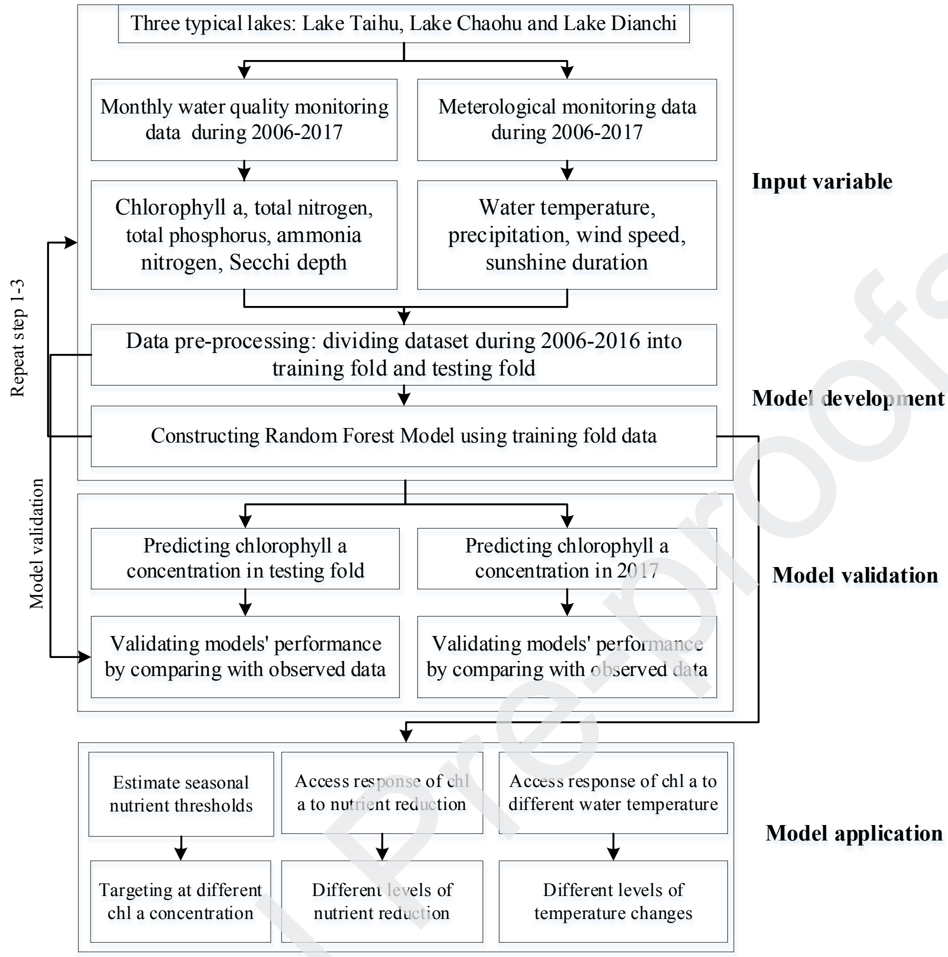


Figure 1. Flowchart of model development and application in predicting nutrient thresholds and response to different nutrient reductions

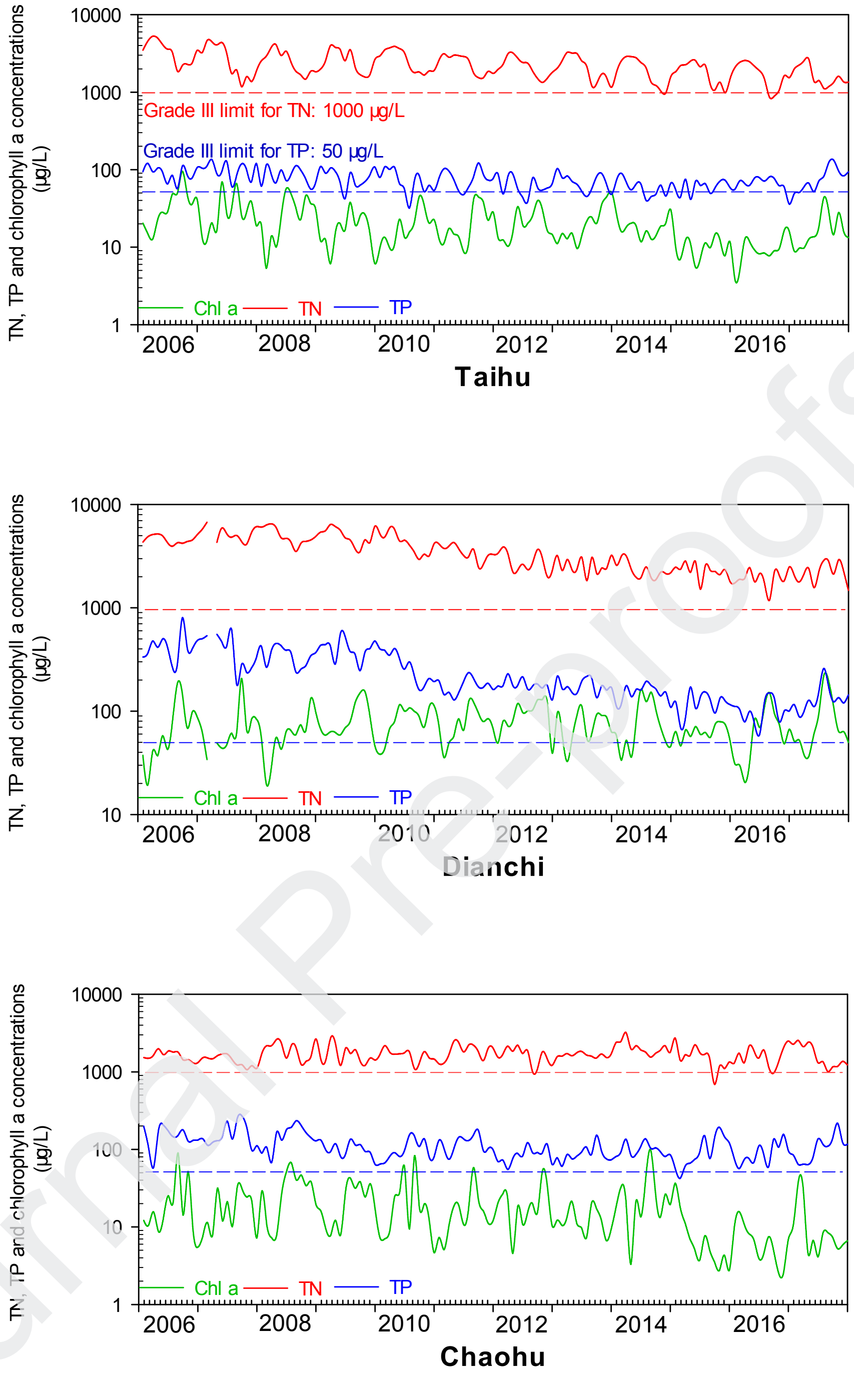
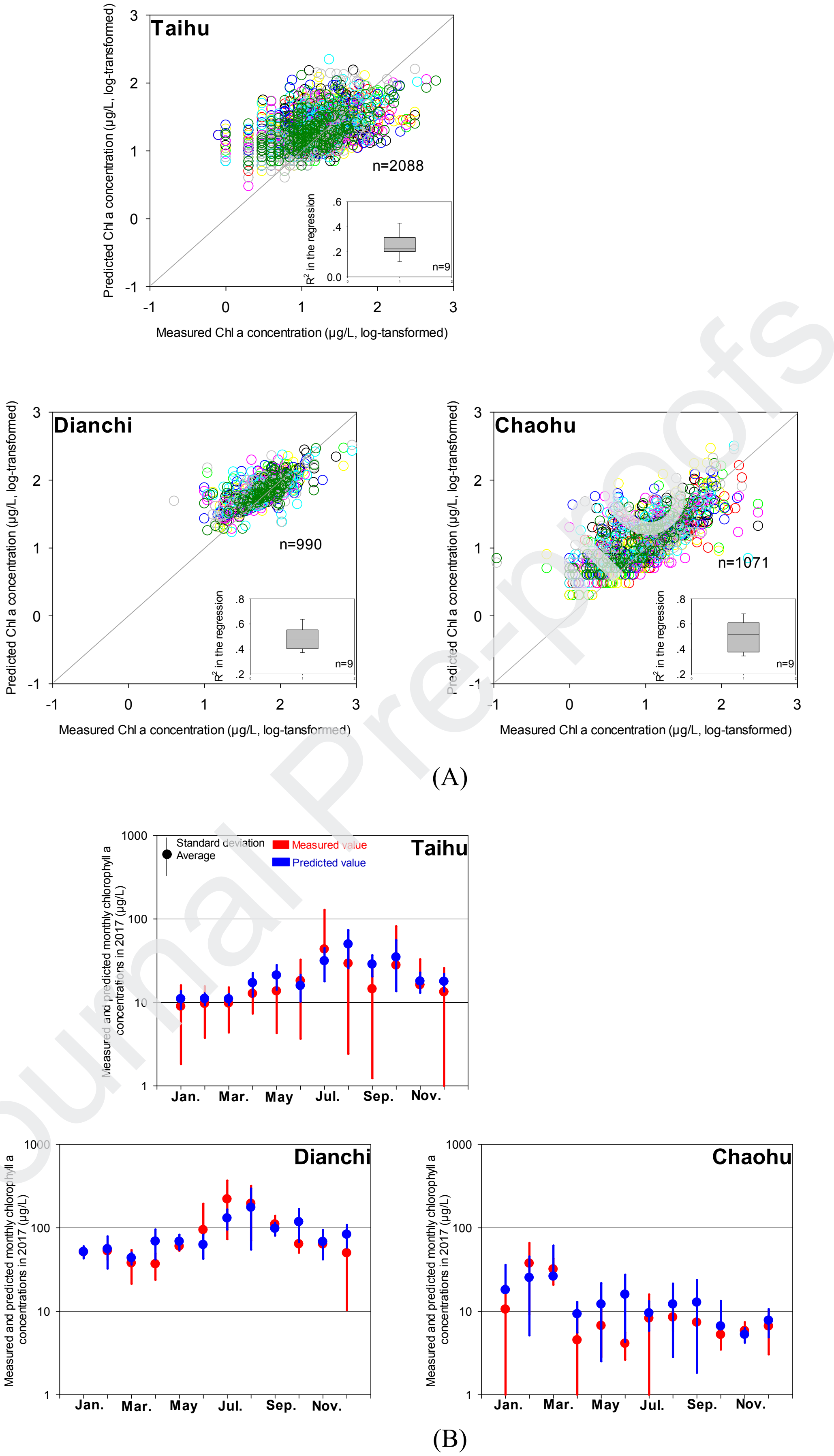
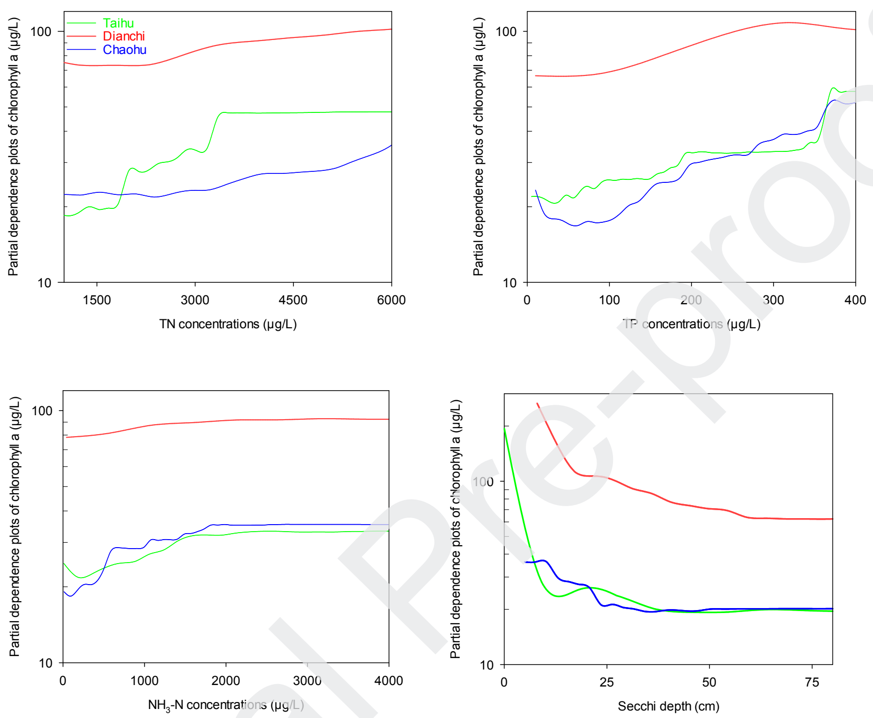


Figure 2. Monthly changes of TN, TP and chlorophyll a concentrations in Lakes Taihu, Dianchi and Chaohu from 2006 to 2017



Taihu
n=2088
n=9
Predicted Chl a concentration (μg/L, log-transformed)
Measured Chl a concentration (μg/L, log-tansformed)
R^2 in the regression
Dianchi
n=990
Chaohu
n=1071
(A)
Standard deviation
Average
Measured value
Predicted value
Measured and predicted monthly chlorophyll a concentrations in 2017 (μg/L)
Jan.
Mar.
May
Jul.
Sep.
Nov.
(B)

Figure 3. Overall cross-validated predictions by the random forest model in Taihu, Chaohu and Dianchi. A. Comparison between predicted chlorophyll a concentration and measured value in the testing dataset; B. Prediction of monthly chlorophyll a concentration in three lakes in 2017.



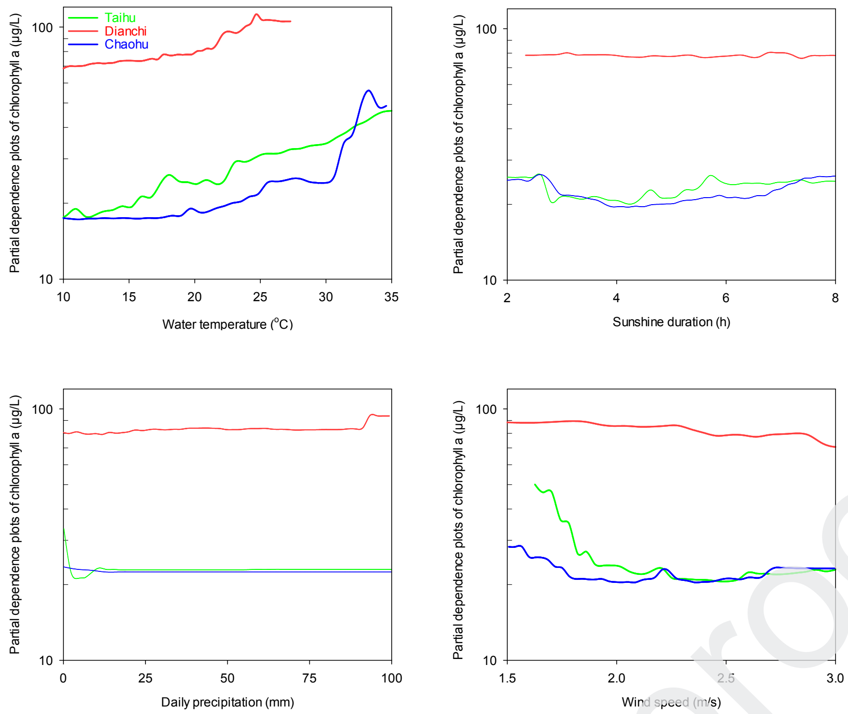
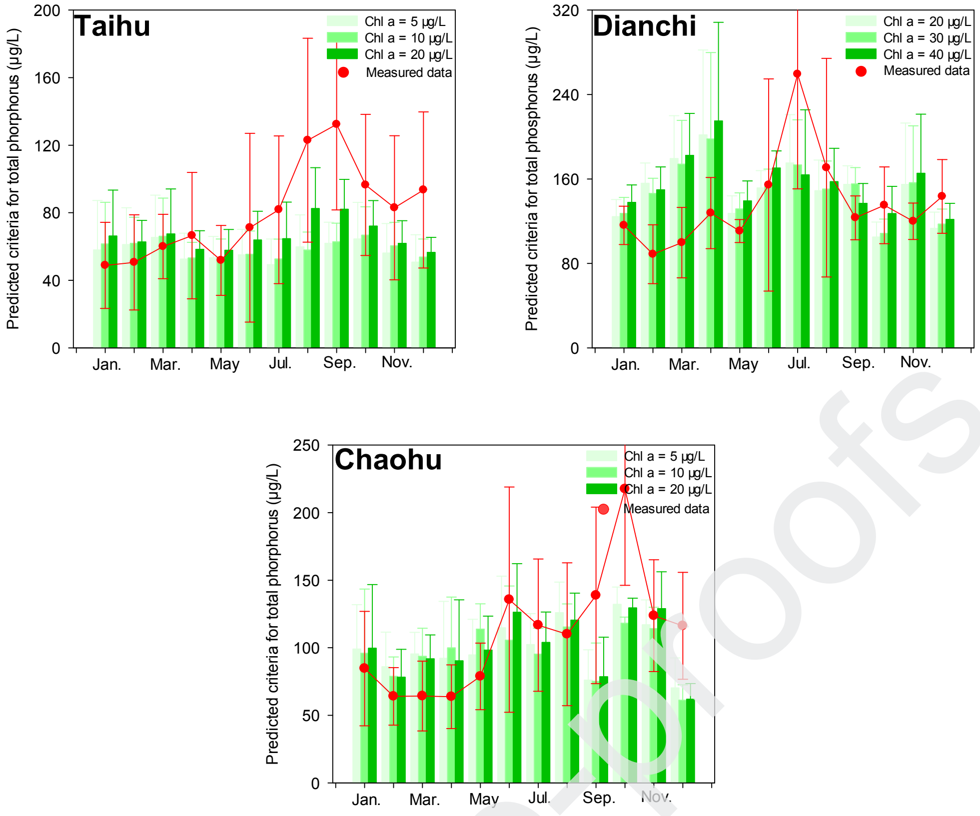
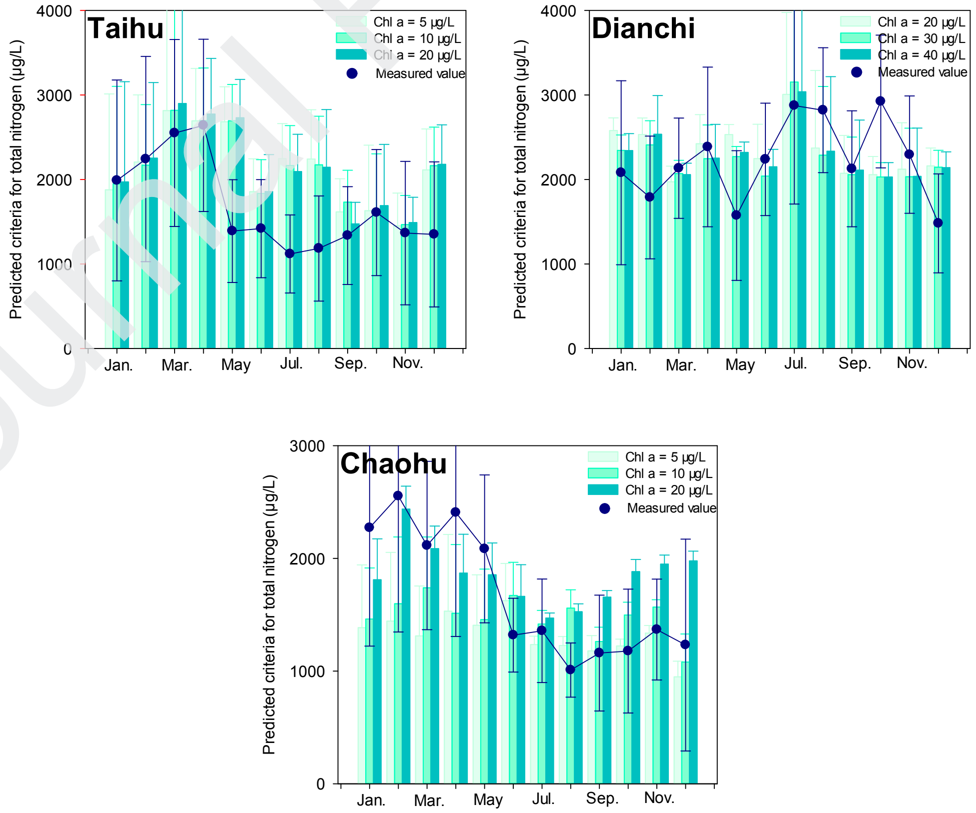


Figure 4. Partial dependence plots of chlorophyll a concentration to different explanatory variables in the random forest model

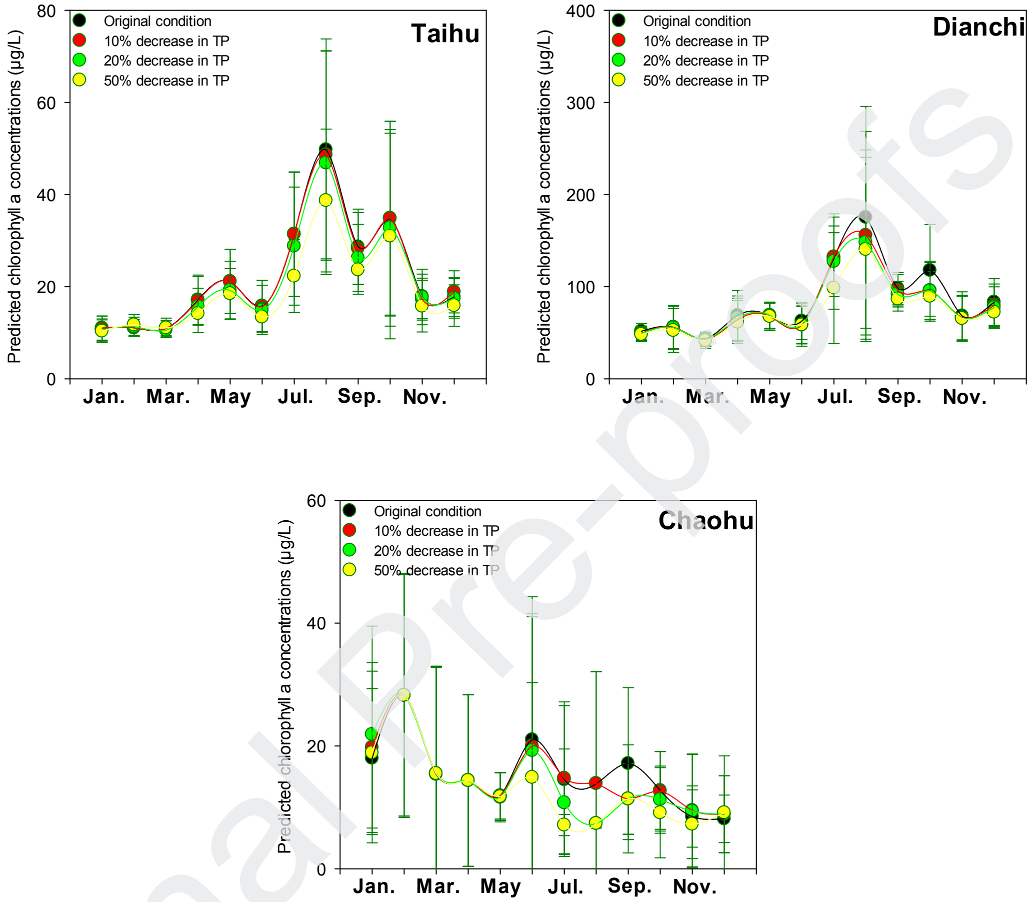

Taihu
Chl a = 5 μg/L
Chl a = 10 μg/L
Chl a = 20 μg/L
Measured data
Predicted criteria for total phorphorus (μg/L)
0
40
80
120
160
200
Jan.
Mar.
May
Jul.
Sep.
Nov.
Dianchi
Chl a = 20 μg/L
Chl a = 30 μg/L
Chl a = 40 μg/L
Measured data
Predicted criteria for total phosphorus (μg/L)
0
80
160
240
320
Jan.
Mar.
May
Jul.
Sep.
Nov.
Chaohu
Chl a = 5 μg/L
Chl a = 10 μg/L
Chl a = 20 μg/L
Measured data
Predicted criteria for total phorphorus (μg/L)
0
50
100
150
200
250
Jan.
Mar.
May
Jul.
Sep.
Nov.

(A)

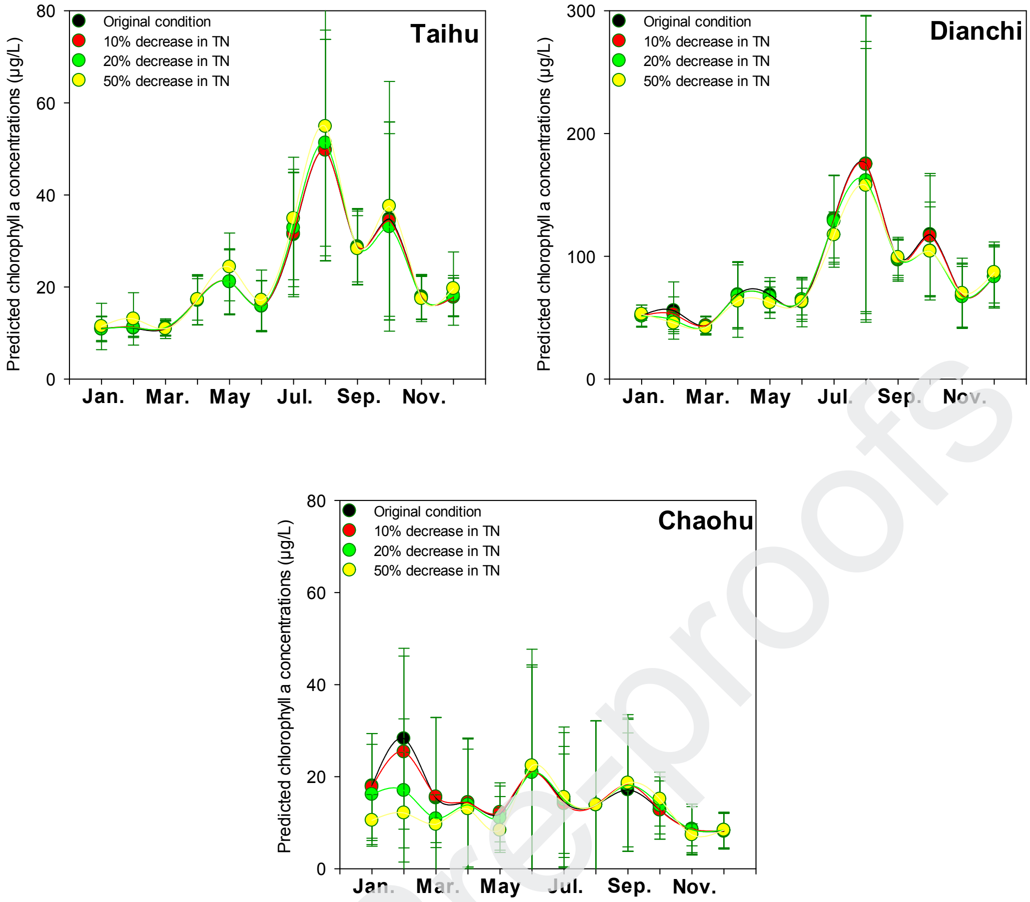

Taihu
Chl a = 5 μg/L
Chl a = 10 μg/L
Chl a = 20 μg/L
Measured value
Predicted criteria for total nitrogen (μg/L)
0
1000
2000
3000
4000
Jan.
Mar.
May
Jul.
Sep.
Nov.
Dianchi
Chl a = 20 μg/L
Chl a = 30 μg/L
Chl a = 40 μg/L
Measured value
Predicted criteria for total nitrogen (μg/L)
0
1000
2000
3000
4000
Jan.
Mar.
May
Jul.
Sep.
Nov.
Chaohu
Chl a = 5 μg/L
Chl a = 10 μg/L
Chl a = 20 μg/L
Measured value
Predicted criteria for total nitrogen (μg/L)
0
1000
2000
3000
Jan.
Mar.
May
Jul.
Sep.
Nov.

(B)

Figure 5. Estimated monthly nutrient thresholds for (A) TP and (B) TN concentration in Lakes Taihu, Dianchi and Chaohu



(A)



(B)

Figure 6. Response of chlorophyll a concentrations to decrease of TP (A) and TN (B) concentrations in the lakes

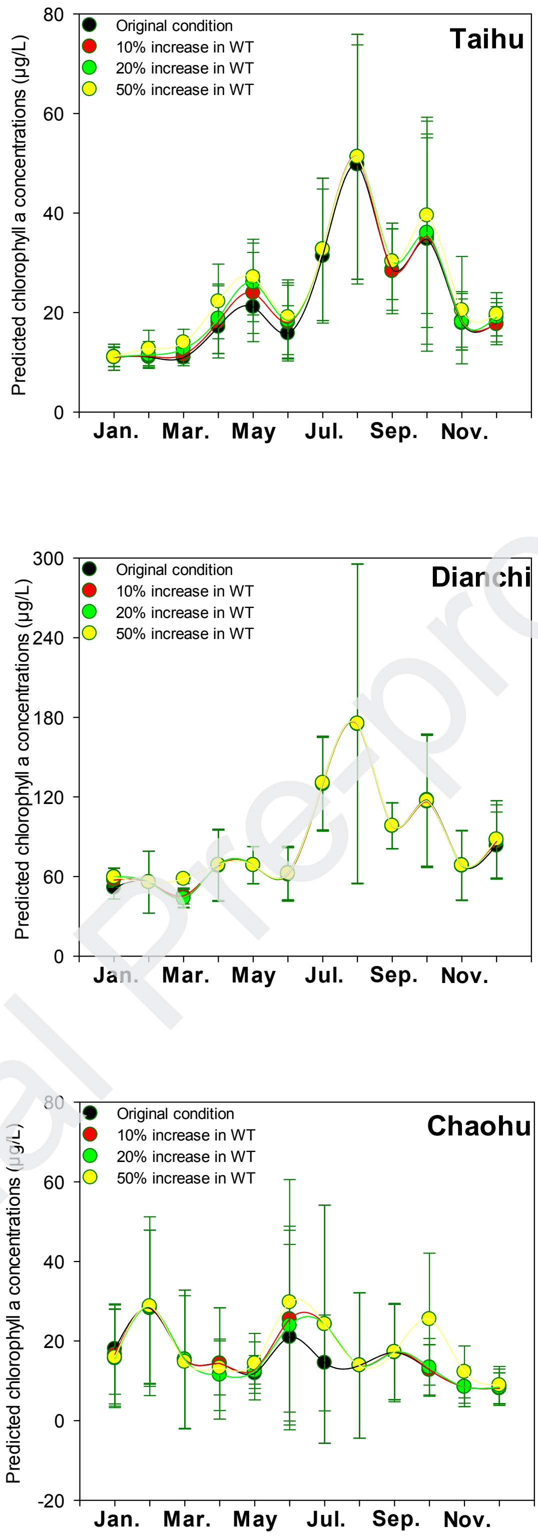


Figure 7. Response of chlorophyll a concentration to the increases of water temperatures in the lakes

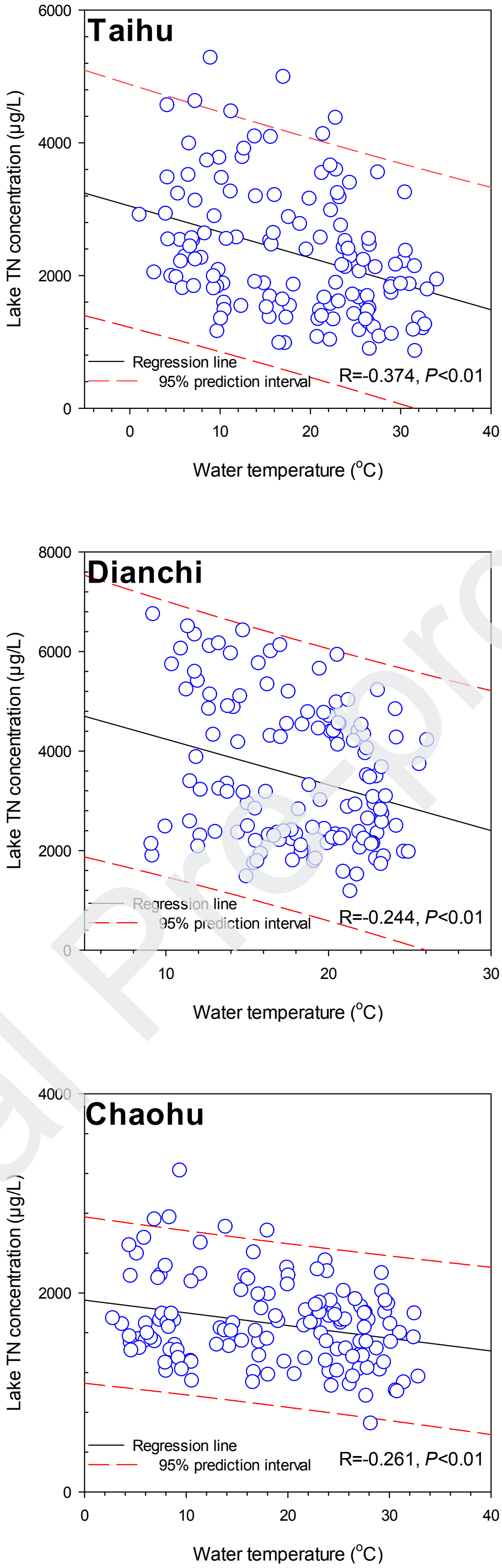


Figure 8. Correlation between water temperature and TN concentrations in Lakes Taihu, Dianchi and Chaohu

Table 1. Basic characteristics of selected lakes and distribution of variables in data set (median and 95% confidence intervals, n=sampling number)

Lake name	Location	Lake area (km^2)	Average depth (m)	Chl a (μg/L)	TN (μg/L)	TP (μg/L)	NH_4^+-N (μg/L)	SD (cm)	
Taihu	E: 120.14°; N: 31.41°	2329	2.2	12 (3-71), n=2800	1920(690-6136), n=2843	60 (20-180), n=2843	110(30-2129), n=2843	30 (10-50), n=2830	19.3(n
Dianchi	E: 102.66°; N: 25.01°	298	5.0	63(18-177), n=1430	2230(1290-1249), n=1429	144(66-1066), n=1429	279(110-8906), n=1428	43(27-76), n=1430	18.7(n
Chaohu	E: 117.38°; N: 31.56°	787	2.6	9.3(1.7-69.0), n=1496	1520(646-3545), n=1496	94(41-275), n=1496	372(108-1460), n=1496	35(15-55), n=1496	19.7(n

(1) All the variables in the data set were monitored monthly enduring from January 2006 to December 2017. (2) The detailed locations of the monitoring sites were provided in Figure S1; (3) Chl a – Chlorophyll a; TN – total nitrogen; TP – total phosphorus; NH_4^+-N – ammonia nitrogen; SD – Secchi depth; WT – water temperature; Rad – daily sunshine duration; Rain – daily precipitation; Wind – wind speed; (4) The data is calculated based on the data set throughout the study period.

Highlights

- Chlorophyll a concentrations predicted by random forest models successfully displayed the seasonal variations.
- Estimated total nutrient thresholds were quite variable and higher in summer than in winter.
- It was more effective to control the TP concentrations in these lakes than the TN concentrations to control algal growth.
- Seasonal variation in nutrient concentrations and environmental drivers should be considered when establishing nutrient criteria.

Declaration of Interest Statement

The authors declare no competing financial interests.

Abstract

Eutrophication and subsequent harmful cyanobacteria blooms are global water quality problems, and identifying the key drivers of water eutrophication and estimating nutrient thresholds for it in waterbodies have long been challenges for water quality managers. Data-intensive machine learning models have been shown to be better able to reveal the nonlinear relationships between variables in the study of complex biotic community dynamics than traditional mechanistic models. In this study, we applied random forest models to long-term datasets from nutrient monitoring and meteorological observations to characterize the relationships between algal growth and different environmental drivers in three eutrophic lakes in China. We further attempted to estimate the season-specific nutrient thresholds in these lakes, and assess the potential decreases in chlorophyll a concentrations that could be achieved through nutrient management. In general, chlorophyll a concentrations predicted by the random forest models were consistent with the values observed in the lakes, and successfully displayed the same seasonal variations. The estimated total nitrogen (TN) and total phosphorus (TP) nutrient thresholds were quite variable among months, and were higher in summer than in winter. To maintain chlorophyll a concentrations below 20 μg/L, the estimated TN thresholds in Lakes Taihu, Dianchi, and Chaohu in August were 2145 ± 683, 2372 ± 918 and 1527 ± 71 μg/L (mean ± standard deviation), respectively, and the corresponding TP thresholds were 82 ± 24, 149 ± 22, and 120 ± 22 μg/L. The modelling results indicated that it was more important to control the TP concentrations in these lakes than the TN concentrations

to control algal growth in summer. In summary, the strong seasonal variation in the estimated nutrient thresholds suggests that a ‘one-size-fits-all’ nutrient control target could overprotect these water bodies. Seasonal variation in nutrient concentrations and environmental drivers should thus be considered when establishing nutrient criteria and setting nutrient control targets.