



Projected response of algal blooms in global lakes to future climatic and land use changes: Machine learning approaches

Jinge Ma^{a,b,d}, Hongtao Duan^{b,c,*}, Cheng Chen^{a,d}, Zhigang Cao^{b,c}, Ming Shen^{b,c},
Tianci Qi^{b,c}, Qiuwen Chen^{a,d,e,**}

^a The National Key Laboratory of Water Disaster Prevention, Nanjing Hydraulic Research Institute, Nanjing 210029, China

^b Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

^c State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China

^d Center for Eco-Environmental Research, Nanjing Hydraulic Research Institute, Nanjing 210029, China

^e Yangtze Institute for Conservation and Green Development, Nanjing 210029, China

ARTICLE INFO

Keywords:

Algal bloom prediction
Global lakes
Climate change
Land use
Machine learning

ABSTRACT

The eutrophication of lakes and the subsequent algal blooms have become significant environmental issues of global concern in recent years. With ongoing global warming and intensifying human activities, water quality trends in lakes worldwide varied significantly, and the trend of algal blooms in the next few decades is unclear. However, there is a lack of comprehensive quantitative research on the future projection of lake algal blooms globally due to the scarcity of long-term algal blooms observational data and the complex nonlinear relationships between algal blooms and their driving factors. We aimed to develop a global projection model to evaluate the future trend in algal bloom occurrences in large lakes under various socio-economic development scenarios. We focused our research on 161 natural lakes worldwide, each exceeding 500 km². The results indicated that the Random Forest model performed best (Overall Accuracy: 0.9697, Kappa: 0.8721) among various machine learning models which were applied in this study. The predicted results showed that, by the end of this century, the number of lakes experiencing algal blooms and the intensity of these blooms will worsen under higher forcing scenarios (SSP370 and SSP585) ($p < 0.05$). In different regions, lakes with increasing algal blooms are mainly distributed in Africa, Asia, and North America, while lakes with decreasing occurrence are primarily found in Europe. Additionally, underdeveloped regions, such as Africa, exhibit greater sensitivity to different SSP scenarios due to high variability in population and economic growth. This study revealed the spatiotemporal distribution of algal blooms in global lakes from 2020 to 2100 and suggested that the intensifying algal blooms due to global warming and human activities may offset the effort of controlling the water quality.

1. Introduction

The algal bloom has emerged as a significant characteristic of global lake water pollution and a pressing environmental issue (Carstensen et al., 2007; McCrackin et al., 2017). Algal blooms are prone to accumulation on the water surface and release toxicants known as algal toxins into the water (Merder et al., 2023; OrihelDiane M. et al., 2012; Shen et al., 2023), leading to mass mortality of aquatic organisms and further deterioration of water quality in lakes (Huisman et al., 2018), posing significant risks to human health and social development

(Jochimsen et al., 1998). For example, in 2007, algal blooms in Lake Taihu resulted in a week-long disruption of clean and safe drinking water supply in Wuxi, China, severely affecting local development and daily life (Qin et al., 2010). In 2020, 330 elephants died in Botswana due to the ingestion of water contaminated with cyanobacteria (Wang et al., 2021).

Recently, some studies showed a global increase in algal blooms in lakes over the past few decades (Fang et al., 2022; Ho et al., 2019; Hou et al., 2022; Ma et al., 2023), but patterns varied by regions (Ma et al., 2020; Wilkinson et al., 2021). Due to the uncertainty of ongoing global

* Corresponding author at: Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, China.

** Corresponding author at: The National Key Laboratory of Water Disaster Prevention, Nanjing Hydraulic Research Institute, Nanjing 210029, China.

E-mail addresses: hduan@niglas.ac.cn (H. Duan), qwchen@nhri.cn (Q. Chen).

<https://doi.org/10.1016/j.watres.2024.122889>

Received 7 July 2024; Received in revised form 21 November 2024; Accepted 28 November 2024

Available online 29 November 2024

0043-1354/© 2024 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

warming and increasing anthropogenic influences, whether algal blooms are expected to become even more prevalent or seldom globally in the coming decades remains unknown (Paerl et al., 2016). Therefore, it is necessary to quantify the impact of future algal blooms on local aquatic ecosystems so that targeted control measures can be taken to prevent health risks caused by algal blooms in hotspots (Hou et al., 2022; Jones et al., 2023). Existing studies have successfully projected lake algal bloom occurrences based on factors such as land use, meteorological environment, and water quality parameters (Janssen et al., 2019; Kakouei et al., 2021; Ralston and Moore, 2020; Tewari et al., 2022). However, there is a lack of comprehensive quantitative research on the future projection of lake algal blooms on a global scale (Jones et al., 2023).

A major obstacle to global practice is the scarcity of long-term observational data at global scales. Previous observation data has often been sparse and unevenly distributed (Kutser, 2004), and there is also a lack of long-term data in areas where algal blooms may occur. The development of remote sensing technology provides possibilities for global observations of water environments, enabling estimation of algal bloom extends from individual lakes to global lakes (Hou et al., 2022; Ma et al., 2023). Another obstacle is the lack of quantification responses to relative driving factors to algal blooms in global lakes, which are necessary for predicting models. Recent studies have used machine learning and other methods to quantify nonlinear complex relationships such as algal blooms and established algal bloom prediction models based on these relationships (Chen et al., 2024; Kakouei et al., 2021; Lin et al., 2023). Furthermore, there is considerable uncertainty in long-term projections, particularly in assessing forecasts related to human activities (Amer et al., 2013), which are also key points in controlling algal blooms (Burford et al., 2020; Huisman et al., 2018; Paerl and Huisman, 2008). Scenario-based future simulations have become an effective method for reducing prediction uncertainty by combining different Representative Concentration Pathways (RCPs) and Shared Socioeconomic Pathways (SSPs) (Woolway et al., 2022). However, how to use machine learning and long-term observation data to establish an algal bloom projection model in global lakes has not been fully explored.

To fill the knowledge above gaps, this study focused on 161 large lakes (with an area greater than 500 km²) globally and aimed to (a) build an algal bloom projection model for these large lakes based on machine learning algorithms, (b) to obtain the algal bloom dataset of these lakes from 2020 to 2100 and (c) to analyze the future spatial-temporal distribution of algal blooms in lakes globally.

2. Materials and methods

2.1. Algal bloom dataset

In this study, we utilised the algal bloom dataset (ABGL2000-2020, Algal Bloom in Global Lakes from 2000 to 2020, more abbreviations see Table S1) of 161 large natural lakes (> 500 km²) globally (Fig. 1), as provided by Ma et al. (2023). ABGL2000-2020 was derived from Moderate-resolution Imaging Spectroradiometer (MODIS) surface reflectance products and involved the following processing steps. Firstly, MODIS surface reflectance products were used to remove lake ice pixels. Then, the Normalized Difference Snow Index (NDSI), Turbid Water Index (TWI), and Cyanobacteria and Macrophytes Index (CMI) were utilised to eliminate lake ice, highly turbid water, and aquatic vegetation areas, respectively. Finally, the algal bloom scum was extracted using a Floating Algae Index (FAI) threshold. The dataset considered images where the percentage of available pixels in a lake exceeded 50 % as valid observations, with a spatial resolution of 250 m.

By quantifying the algal bloom pixels in every lake, we acquired a daily-scale algal bloom record for the 161 large natural lakes worldwide spanning from 2000 to 2020. Algal blooms were identified based on the criterion of the algal bloom area surpassing 5 % of the lake's total surface area. The data was converted to the monthly scale based on the daily algal bloom dataset that was obtained to increase the adequate sample size. According to the maximum area of algal bloom for each month, we determined whether an algal bloom occurred during that month, thus creating a monthly algal bloom dataset.

2.2. Driving factor dataset

Lake Morphology: The average depth and surface area of study lakes were obtained from the HydroLAKES database. The Dynamic Sediment Ratio (DSR) is the square root of the surface area divided by the average depth. It determines the sensitivity of lakes to sediment resuspension caused by wind-induced waves (Jansson, 1985; Shen et al., 2022a). A higher DSR value indicates lakes are more prone to sediment resuspension caused by wind disturbance, providing nutrients for algal growth.

Meteorological Data. The daily temperature, wind speed, precipitation, and atmospheric pressure data for each lake were obtained from the ERA5-Land dataset, the fifth-generation reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF), covering the period from 2000 to 2020. This dataset covers the global

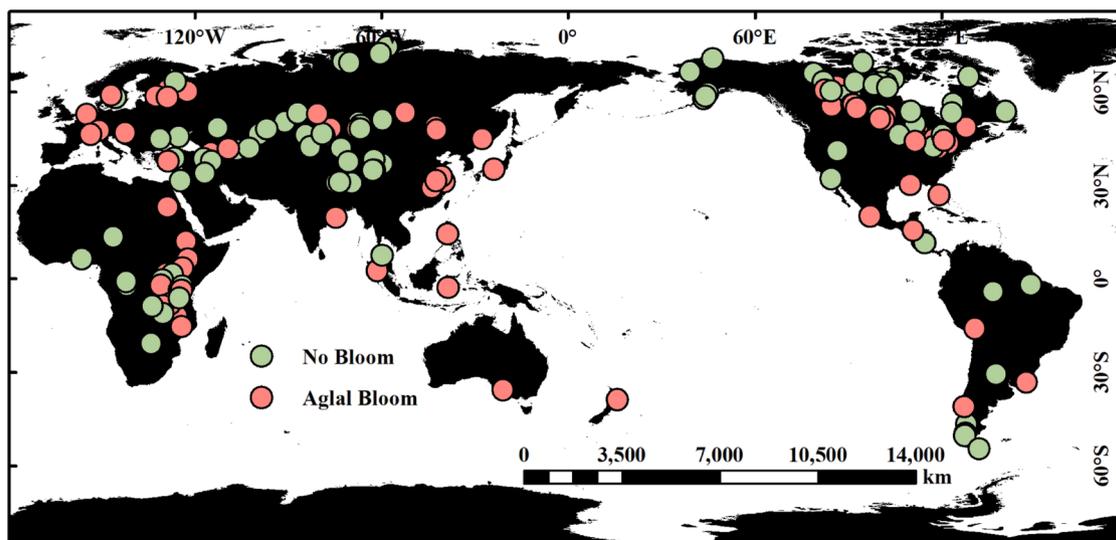


Fig. 1. Distribution of study lakes with and without algal blooms based on the ABGL2000-2020 dataset.

land surface with a temporal resolution of hourly data and a spatial resolution of 0.1° (Muñoz Sabater, 2019). Using the daily data, we further calculated the annual average temperature, average wind speed, average precipitation, average atmospheric pressure data, and the overall average values for the entire study period. Temperature, wind speed, and atmospheric pressure data were clipped using the lake boundaries, while precipitation data were clipped using the basin boundaries.

Land Use/Land Cover. The land use/land cover data within each lake's basin for the study period were obtained from the MCD12Q1 dataset. MCD12 is a series of products derived from the MODIS images, providing a global land cover classification distribution (Sulla-Menashe and Friedl, 2019). The MCD12Q1 provides yearly land cover classification with a resolution of 500 m, which classifies the global land into 17 categories, including forests, grasslands, wetlands, croplands, and urban areas. We aggregated the classification data from MCD12Q1 into five land use types: forests, croplands, urban areas, wetlands, and grasslands.

Population Density. The yearly population density was obtained from the LandScan dataset for the years 2000 to 2020 (Dobson et al., 2000). LandScan is a high-resolution global population distribution dataset with a resolution of 1000 m, offering population counts and densities (individuals per square kilometer) at each pixel. Population density served as a proxy metric for gauging the intensity of human activity within the respective regions.

Basin Boundary: The basin boundaries for study lakes were obtained from the HydroBASINS dataset (Lehner and Grill, 2013). For land use data, we calculated the coverage percentage of each land use type within each lake's basin boundary. It should be noted that the HydroBASINS usually excludes the water body within the lake's basin boundary and only calculates the coverage percentage of the remaining land use types. Additionally, for other datasets pertaining to human activity, we determined the mean values within each lake's basin boundary. For more comprehensive insights into the driving factors considered in this study, please refer to Table S2.

2.3. Future forcing data

Meteorological data. We downloaded the global climate model data of climate forcings provided by the ISIMIP3b module as future climate changes (<https://data.isimip.org/search/tree/ISIMIP3b/InputData/climate/>). We selected downscaled output meteorological data from five representative global climate models participating in ISIMIP. These models (Table S3) include the GFDL Earth System Model version 4 (GFDL-ESM4) proposed by NOAA, the latest version of the climate model from the Institut Pierre Simon Laplace (IPSL-CM6A-LR) in France, the Max Planck Institute for Meteorology Earth System Model (MPI-ESM1-2-HR) in Germany, the Meteorological Research Institute Earth System Model 2.0 (MRI-ESM2-0) in Japan, and the UK Earth System Model (UKESM1-0-LL) proposed by the Natural Environment Research Council (NERC) in the United Kingdom.

To standardize the temporal scope of all model data, we extracted information falling within the timeframe of 2020-2100 from each model as future climate input data. Subsequently, we computed the monthly averages as well as the inter-model averages for each dataset. From the five models, we isolated climate-forcing data encompassing temperature, wind speed, precipitation, and air pressure. Temperature values are represented in Kelvin (K), wind speed in meters per second (m/s), precipitation in kilograms per second (kg/s), and air pressure in Pascals (Pa). To enhance comparability, we converted temperature units to Celsius ($^\circ\text{C}$) and precipitation units to millimeters per month (mm/Month). We calculated the mean values for each meteorological variable within the lake boundaries for each lake. The temperature, wind speed, and air pressure data were cropped using the lake boundaries to ensure coherence between the meteorological variables interacting with lakes. Conversely, the precipitation data was adjusted using the basin

boundaries for precision in analysis.

Land use/land cover. We obtained future land use/land cover datasets from the Northwest Pacific National Laboratory in the United States (<https://doi.org/10.25584/data.2020-07.1357/1644253>). The data cover the period from 2015 to 2100, with a temporal resolution of 5 years and a spatial resolution of 0.05° . It includes 15 coupled scenarios of Shared Socioeconomic Pathways (SSPs) and Representative Concentration Pathways (RCPs) (Chen et al., 2020). The dataset encompasses a total of 32 land use types, including forests, croplands, grasslands, bare land, urban land, shrubs, and more. For our study, we specifically selected the SSP-RCP coupled scenarios (SSP1-RCP2.6, SSP3-RCP7.0, and SSP5-RCP8.5) that align with the meteorological input data. We condensed the expansive dataset into five primary categories: forests, urban land, bare land, croplands, and grasslands. Subsequently, utilizing the lake basin boundaries delineated by HydroBASINS, we computed the coverage percentage of each land use/land cover type within the basin boundary of every lake. This analysis aids in understanding the spatial distribution and composition of land use within the vicinity of each lake.

Population density. The population density grid dataset for future projections from the School of Architecture at Tsinghua University (<https://doi.org/10.6084/m9.figshare.19608594.v2>). The data cover the period from 2020 to 2100, with a temporal resolution of 5 years and a spatial resolution of 1000 m, which includes five SSP scenarios (SSP1-5) (Wang et al., 2022). We calculated the mean population density within the basin boundary of each lake in each scenario.

Gross Domestic Product (GDP). The GDP dataset for future projections originates from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (<http://doi.org/10.5281/zenodo.4350027/>). This dataset spans from 2000 to 2100, featuring a temporal resolution of 5 years and a spatial resolution of 1000 m. It encompasses projections across five Shared Socioeconomic Pathways (SSP) scenarios (Wang and Sun, 2021). We calculated the mean GDP within the basin boundary of each lake in each scenario. For detailed information on future forcing data in this study, see Table S4.

2.4. Algal bloom projection model

Building the training and validation sets. The process of modeling and simulating algal blooms in lakes is depicted in Fig. 2. The algal bloom dataset (ABGL2000-2020) and driving factor dataset were divided into monthly training and validation datasets (Part I in Fig. 2). We first selected data on meteorological factors, human activity factors, and lake morphological characteristics to construct the global lake algal bloom projection model. The previous study and literature review (Huang et al., 2014; Huisman et al., 2018; Reichwaldt and Ghadouani, 2012) selected atmospheric pressure, temperature, precipitation, and wind speed as meteorological forcing factors. Considering the driving mechanisms of algae growth and bloom formation, the cumulative temperature and precipitation of the previous 1–3 months were calculated to provide more meteorological information for accurate algal bloom prediction. We also selected cropland coverage, GDP representing economic activity intensity, and population density as indicators of human activity intensity. Meanwhile, average lake depth, surface area, DSR, and lake retention time were chosen as indicators of lake morphological characteristics.

The original meteorological data was initially on a daily scale. This study aggregated and calculated the original data into monthly values. Population density and cropland coverage were measured annually for human activity data, while GDP data were collected every five years. We assumed that the data changed continuously between two consecutive records and distributed the values evenly each month to obtain monthly data sequences. Following these principles, the monthly algal bloom and driving force datasets for 161 lakes from 2000 to 2020 were processed, resulting in 38,962 records (Fig. S1). We randomly allocated 70 % of the data as the training set (DS-AB-Train), consisting of 27,273 records for the model training. In contrast, the remaining 11,689 records were

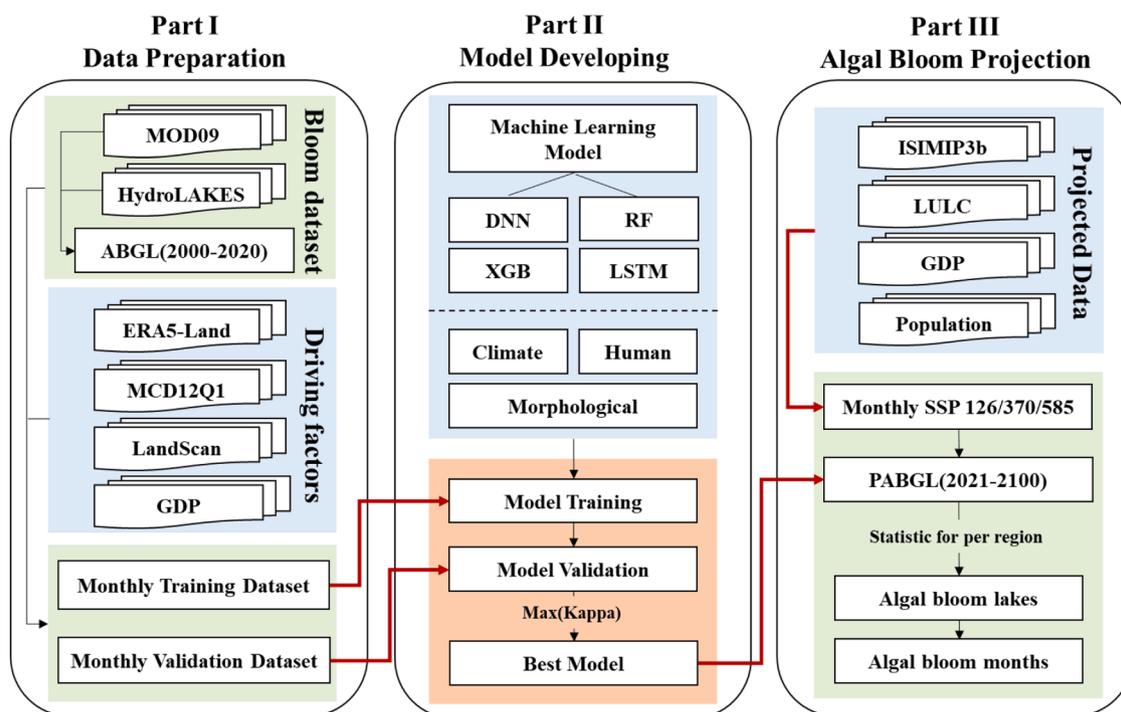


Fig. 2. Flowchart for developing the PABGL dataset. The flowchart is divided into three parts (Part I - Part III): data preparation, model development, and algal bloom projection.

assigned to the validation dataset (DS-AB-Val).

Construction of the algal bloom prediction model. We selected four machine learning algorithms (DNN, RF, XGBoost, and LSTM) with 16 experiments to construct the algal bloom prediction model (Part II in Fig. 2). Considering the complex nonlinear relationship between algal blooms and driving forces, we designed 16 experiments to determine the optimal input parameter variables and machine learning models. There are four parameter combinations used (Table S5): meteorological parameters + human activity parameters + lake morphological parameters (Model 1), meteorological parameters + human activity parameters (Model 2), meteorological parameters + lake morphological parameters (Model 3), and human activity parameters + lake morphological parameters (Model 4). The optimal hyperparameters for each model were obtained through grid search (Table S6) (Cao et al., 2020; Shen et al., 2022b), and the input datasets for different parameter combinations remained consistent across the same model (Table S7).

Validation and uncertainty analysis. The 5-fold method for cross-validation was employed, evaluating model performance using five randomly selected training and validation sets (based on the Kappa coefficient, OA, and other parameters on the validation set). The Monte Carlo method was also used to estimate the uncertainty of machine learning models and the sensitivity of the model to the input variable errors. A Monte Carlo simulation was performed by repeatedly sampling from the input variable distributions (1000 iterations). For each sample, we calculated uncertainty metrics such as root-mean-square error (RMSE), standard deviation (STD), and confidence intervals. It was assumed that the relative errors of the input variables followed a normal distribution, and the relative errors were the same as those reported in the literature (Bonshoms et al., 2022; Li et al., 2017; Wang and Sun, 2021; Wang et al., 2022). The five parameters with the highest degree of influence on the model were selected as the analysis objects. Assuming that the upper limits of the relative errors of these products were 40 %, 20 %, 40 %, and 20 %, respectively, the relative errors were randomly assigned 1000 times and substituted into the estimation model. Parameters that were less than 0 after the assigned errors were reassigned to ensure the model's regular operation.

Simulation dataset for algal bloom projection. The Projected algal bloom dataset used the best model and projected driving factor dataset (Part III in Fig. 2). The best algorithm simulated datasets for different RCP and SSP scenarios from 2020 to 2100. Assuming that the driving mechanisms remain unchanged during the simulation period, a monthly dataset of algal blooms in 161 large lakes ($>500 \text{ km}^2$) globally was simulated (referred to as PABGL2020-2100, Projected Algal Bloom in Global Lakes from 2020 to 2100). Based on the statistical results of different GCMs, the maximum, minimum, and average values of lakes with algal blooms were calculated under various scenarios. The average number of months per year in which algal blooms occurred in each region was used to indicate algal bloom intensity (Table S3). Statistical analyses were conducted on algal bloom occurrences in lakes across continents and climate regions from 2020 to 2100. Under the same SSP scenario, the average values of different model results were calculated and presented for various periods, such as 2000-2020, 2023-2040, 2041-2060, 2061-2080, and 2081-2100.

3. Results

3.1. Evaluation of the algal bloom prediction models

From the perspective of the accuracy of different input models (Table S8), Model 1 achieved the highest accuracy for the machine learning models (OA > 0.8757 , Kappa > 0.5626). Compared to Model 2 (OA > 0.5799 , Kappa > 0.2044), Model 3 (OA > 0.3168 , Kappa > 0.062), and Model 4 (OA > 0.8624 , Kappa > 0.0011), Model 1 significantly improved the prediction accuracy. Among the different machine learning algorithms, the Random Forest Classification model (RFC, OA > 0.9568 , Kappa > 0.8056) achieved the highest accuracy, followed by the eXtreme Gradient Boosting model (XGB, OA > 0.9269 , Kappa > 0.6152). The Deep Neural Network model (DNN, OA > 0.8554 , Kappa > 0.062) and the Long Short-Term Memory model (LSTM, OA > 0.3168 , Kappa > 0.011) had relatively lower accuracies.

Specifically, for Model 1 (Table 1), the RFC1 algorithm exhibited the highest accuracy (OA = 0.9697, Kappa = 0.8721, Precision > 0.8762 ,

Table 1
Statistical results for the validation of different machine learning approaches in Model 1.

Model	Classification	Precision	Recall	F1-score	OA	Kappa
RFC1	NoBloom	0.9843	0.9808	0.9825	0.9697	0.8721
	Bloom	0.8762	0.8967	0.8863		
XGB1	NoBloom	0.9613	0.9849	0.9730	0.9525	0.7770
	Bloom	0.8814	0.7388	0.8038		
DNN1	NoBloom	0.8808	0.9910	0.9327	0.8757	0.5626
	Bloom	0.6583	0.6147	0.6953		
LSTM1	NoBloom	0.8108	0.7692	0.7895	0.7838	0.5676
	Bloom	0.7568	0.8000	0.7778		

Recall > 0.8967, F1-score > 0.8863), surpassing the XGB1 algorithm (OA = 0.9525, Kappa = 0.7770, Precision > 0.8814, Recall > 0.7388, F1-score > 0.8038), DNN1 algorithm (OA = 0.8757, Kappa = 0.5626, Precision > 0.6583, Recall > 0.6147, F1-score > 0.6953), and LSTM1 algorithm (OA = 0.7838, Kappa = 0.5676, Precision > 0.7568, Recall > 0.8000, F1-score > 0.7778). The RFC1 algorithm exhibited higher accuracy in predicting non-algal bloom samples (Precision = 0.9843, Recall = 0.9808, F1-score = 0.9825) than in predicting algal bloom samples (Precision = 0.8762, Recall = 0.8967, F1-score = 0.8863). We also trained the RFC1 model with data from 2000 to 2015 and validated it with untouched data from 2016 to 2020 from several lakes on different continents. The results (Table S9 and Fig. S3) showed that although the accuracy (OA = 0.9430, Kappa = 0.7617) is lower than that of the test set (OA = 0.9697, Kappa = 0.8721), it is still acceptable for the goal of predicting long-term series in the future. In addition, the accuracy of the model varies in lakes in different regions, which may be related to the number of input samples in different regions and different types of lakes.

From the perspective of uncertainty (Table S10), the differences between various machine learning models are more significant than the variances of different input parameter types. The uncertainties of XGB (RMSE: 0.1926-0.3626, STD: 0.0119-0.0182), DNN (RMSE: 0.2990-0.3626, STD: 0.0119-0.0182), and LSTM (RMSE: 0.3097-0.3351, STD: 0.1558-0.2285) are notably higher than RF (RMSE: 0.0671-0.3626, STD: 0.0119-0.0182). Across different combinations of input parameters for RFC1, the variations in RMSE and STD are minimal. Further analysis of the sensitivity of RFC1 based on input variables (Tables S11 and S12) reveals that as each input variable or a blend of multiple input variables increases incrementally, the bias remains lower than RFC1’s inherent bias (RMSE = 0.102, STD = 0.0219) until reaching the maximum error threshold. This suggested that the errors associated with input variables exerted a limited overall impact on estimation outcomes, indicating that RFC1 could consistently deliver stable and dependable results.

3.2. Trend of numbers of lakes with algal blooms

We obtained the variation of algal bloom occurrences in large lakes globally from 2020 to 2100 under different RCP-SSP scenarios and GCMs. To facilitate analysis, we aggregated the data from a yearly scale (Figs. S4 and S5) to a 20-year scale. The number of lakes with algal

blooms in large lakes globally is expected to show significant differences across different SSP scenarios (Fig. 3). Despite varying outcomes with different GCMs across all scenarios, the average values increase notably ($p < 0.01$). Among the distinct SSP scenarios, the rising trajectory in the SSP585 ($\beta = 0.06, p < 0.001$) is expected to surpass that in the SSP370 ($\beta = 0.02, p < 0.001$), while the SSP126 displays the smallest escalation ($\beta = 0.01, p < 0.01$). Moreover, within each SSP scenario, the variation in the maximum number of affected lakes from 2020 to 2100 is insignificant, while the variation in the minimum number of affected lakes closely followed the trend of the mean values.

Regarding different input models (Fig. 4), with the exception of the GFDL-ESM4 model, the trends in algal blooms across various SSP scenarios are expected to show insignificant changes ($p > 0.1$). Nevertheless, in most models, the number of lakes impacted by algal blooms is projected to increase substantially ($p < 0.01$) across all three SSP scenarios. The IPSL-CM6A-LR model is expected to exhibit the highest growth rate in the number of affected lakes under both the SSP370 scenario ($\beta = 0.074, p < 0.001$) and the SSP585 scenario ($\beta = 0.085, p < 0.001$). The MPI-ESM1-2-HR model is prone to show the highest growth rate under the SSP126 scenario ($\beta = 0.019, p < 0.001$).

The response to SSP scenarios varies across different regions (Fig. 5). Except for GFDL-ESM4, Africa is expected to show a high sensitivity to the SSP scenarios in terms of the number of lakes experiencing algal blooms from 2020 to 2100. In most models, the number of lakes with algal blooms is prone to be the highest under the SSP585 scenario. Taking average value as an example, the number of lakes with algal blooms is projected to be SSP585 (10 lakes) > SSP370 (8 lakes) > SSP126 (5 lakes). Under the SSP585 scenario, Africa is prone to experience the same number of lakes with algal blooms as Asia (10 lakes), according to both the GFDL-ESM4 and UKESM1-0-LL models, positioning Africa as another global hotspot for algal blooms. Other continents are expected to show no significant changes in response to SSP scenarios across different models.

3.3. Trend of months with algal blooms in global lakes

In different continents (Fig. 6), the trend in the average bloom month varied significantly among SSP scenarios, with more significant trends observed in the SSP370 and SSP585 scenarios. The trend in the bloom months in lakes in Africa didn’t show a substantial change ($p > 0.05$), exhibiting fluctuating variations throughout the study period. In Asia, the average bloom months in lakes are expected to increase from 2000 to 2100 in the SSP370 ($\beta = 0.07, p < 0.05$) and SSP585 ($\beta = 0.05, p < 0.05$) scenarios. The trend is projected to be highest in the SSP370 scenario, where the average bloom month by the end of the century (1.03 months) is higher than in the SSP585 scenario (0.99 months). In Europe, there are substantial differences in the average bloom month among different SSP scenarios, with the average bloom month in the SSP585 by the end of the century (0.95 months) higher than SSP370 (0.63 months) and SSP126 (0.46 months). North America is expected to have minor differences among different SSP scenarios, with noticeable

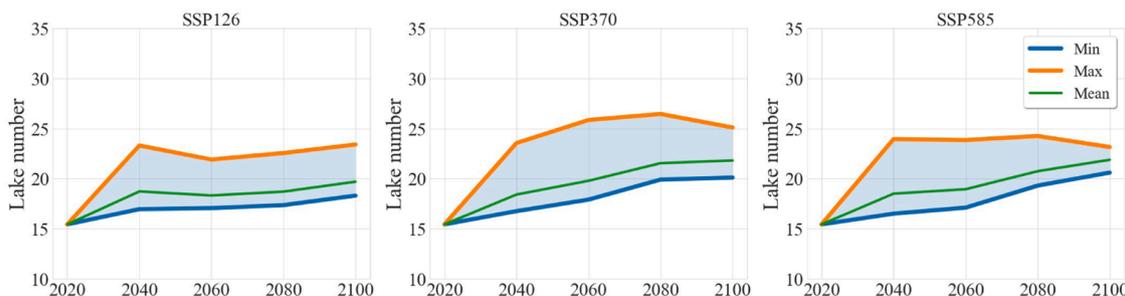


Fig. 3. Range and trends in the number of lakes with algal blooms in a year under different SSP scenarios from 2020 to 2100. The orange line represents the maximum lake number, while the blue line represents the minimum lake number. The light blue area represents the variation range of lake number.

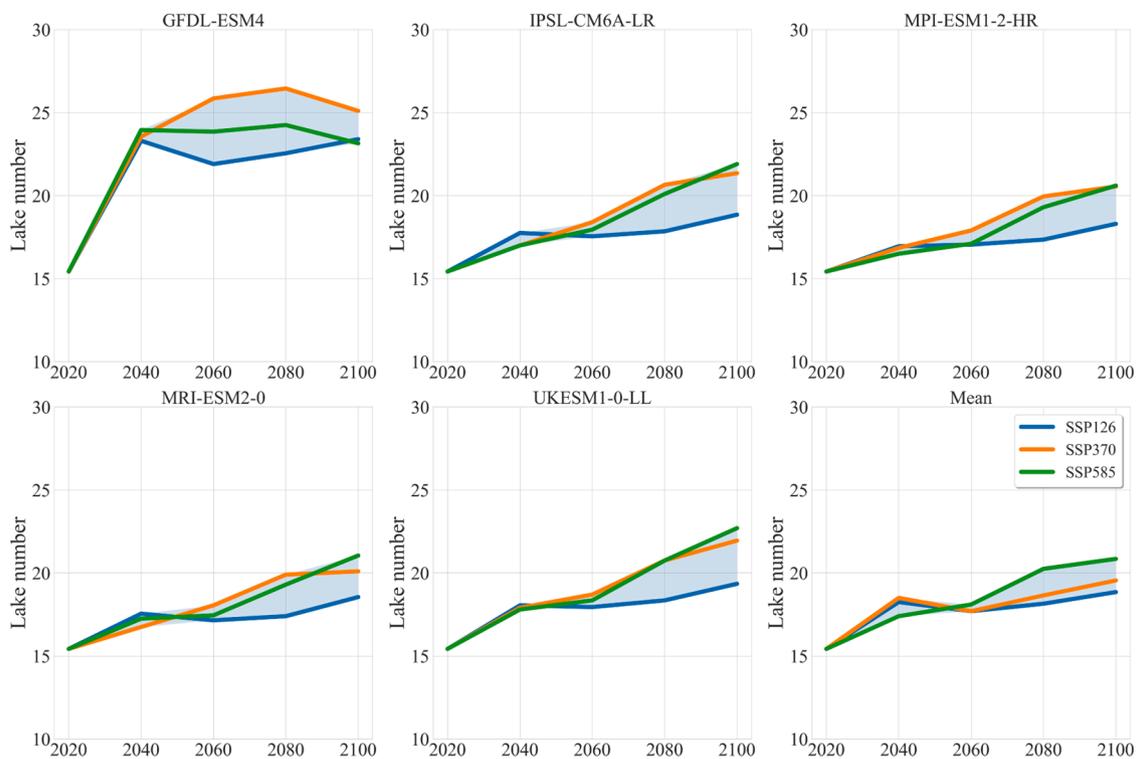


Fig. 4. Range and trends in the number of lakes with algal blooms in a year under different GCMs from 2020 to 2100. The blue line represents the lake number of SSP126, the orange line represents the lake number of SSP370, and the green line represents the lake number of SSP585. The light blue area represents the variation range of lake number.

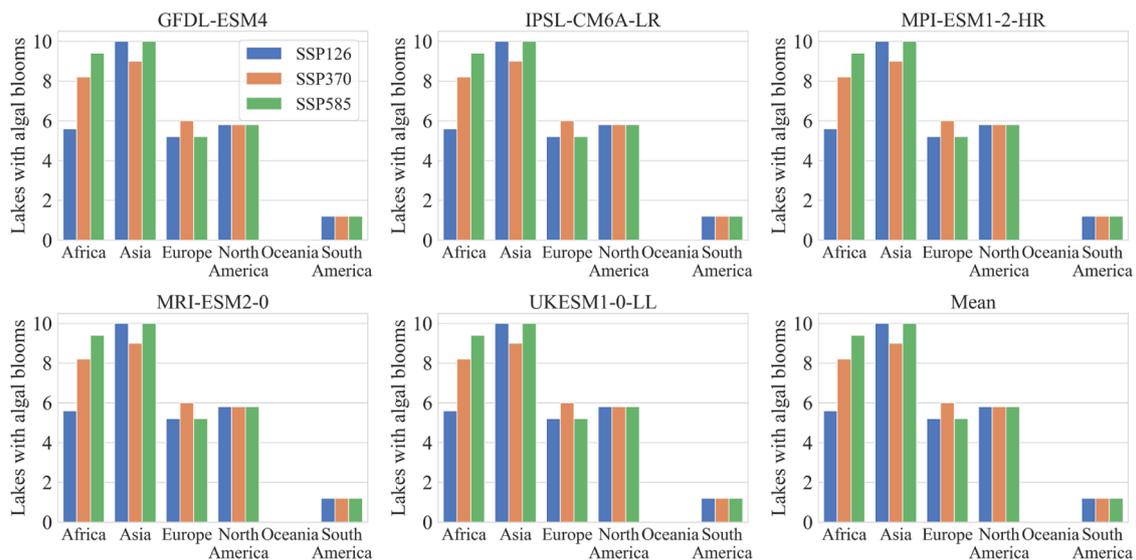


Fig. 5. Number of lakes with algal blooms in different continents from 2020 to 2100 under different SSP scenarios.

differences occurring after 2061-2080. Specifically, in North America, the bloom months in lakes are projected to increase significantly in the SSP370 ($\beta = 0.04, p < 0.05$) and SSP585 ($\beta = 0.08, p < 0.05$) scenarios. In South America, the bloom months are expected to show an overall trend of first decreasing and then increasing, but the trend is insignificant ($p > 0.05$). Oceania will maintain a low level of algal blooms throughout this century.

The spatial distribution of algal bloom months in lakes was illustrated in different scenarios (Fig. 7). Under the SSP126 scenario, lakes in Africa, Asia, North America, and Europe exhibited improvements in the bloom month (i.e., a decreasing trend). Europe is expected to have the

highest number of improved lakes (5 lakes), mainly concentrated in high latitudes. Lakes with worsening algal bloom status are more prevalent in Africa (4 lakes), Asia (5 lakes), and Europe (3 lakes). Under the SSP370 scenario, the bloom months in lakes are projected to worsen in Africa (4 lakes), Asia (4 lakes), Europe (4 lakes), and North America (1 lake). In contrast, under the SSP585 scenario, the bloom months in lakes are prone to worsen in Africa (6 lakes), Asia (5 lakes), Europe (6 lakes), North America (4 lakes), and Oceania (1 lake). The spatial variation of the trend in algal bloom months indicates that regions with higher latitudes and colder weather are prone to experience improvements in the bloom months, while regions with warmer weather are expected to

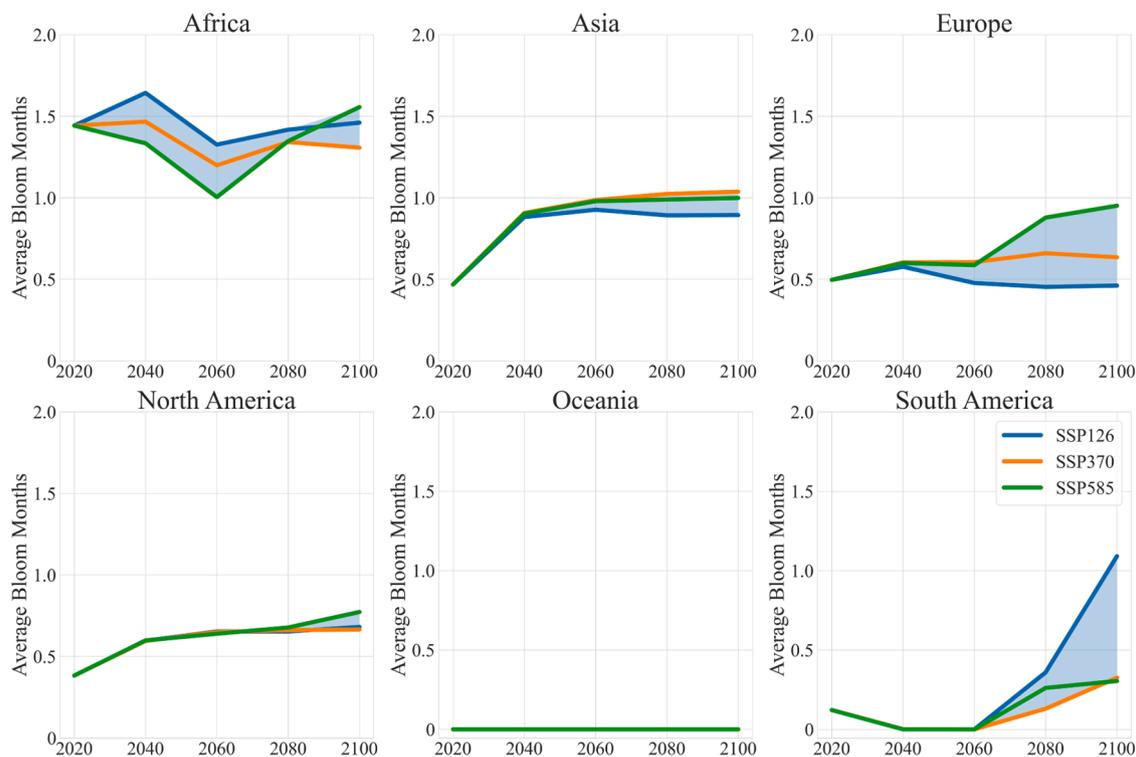


Fig. 6. Trends of average months with algal blooms in different continents from 2000 to 2100 under different SSP scenarios. The blue line represents the bloom month of SSP126, the orange line represents the bloom month of SSP370, and the green line represents the bloom month of SSP585. The light blue area represents the variation range of bloom month.

experience worsening algal blooms.

4. Discussion

4.1. Implications for water management in global lakes

Our results indicated that by the end of this century, both the frequency and intensity of algal blooms in large lakes worldwide are expected to increase under the SSP370 and SSP585 scenarios. This finding aligns with several regional-scale studies, all of which show that the risk of algal blooms in lakes will be higher under high-forcing scenarios (Guan et al., 2022; Kakouei et al., 2021; Zhou et al., 2022). On the one hand, this may be related to anticipated climate change. Higher average temperatures and frequent extreme heat waves under high-forcing scenarios will increase lake water temperatures (Piccolroaz et al., 2024; Tong et al., 2023), leading to earlier lake stratification (Mullin et al., 2020) and earlier algae growth. Higher temperatures are prone to nitrogen limitation in lakes, which is conducive to increasing the dominance of certain non-nitrogen-fixing cyanobacterial species because these species tend to grow best under extreme conditions (Chen et al., 2016; Elliott, 2012). On the other hand, changes in the intensity and nature of human activities, such as land use and population density, have also significantly influenced the input of nutrients from the basin to the lakes (Couture et al., 2014; Pelletier et al., 2015). Therefore, higher population density and urban development under high-forcing scenarios will lead to an increase in the level of nutrient input into lakes worldwide (Chen et al., 2020; Tellman et al., 2021; Wang et al., 2023; 2022; Zang et al., 2021), providing more material basis for algae growth and promoting algal blooms.

Among different regions, the highest increase in the number and intensity of algal blooms is projected in Africa and tropical areas, which are also more sensitive to different scenarios. They are more likely to be impacted by rapid economic growth, which aligns with recent predictions of deteriorating surface water quality in underdeveloped

regions such as Africa (Jones et al., 2023). Furthermore, Africa's increasing population density and GDP during the study period may contribute the most to the trend of algal blooms. The growing water demand, population pressure, and continuous increase in agricultural production and nutrient inputs may motivate algal blooms in African lakes and similar regions. At the same time, Africa is facing a severe shortage of clean water (Zhao et al., 2023), so decisive measures must be taken to ensure the achievement of the United Nations SDG6 goals.

However, although there have been few precedents of water quality improvement in certain regions, most improved lakes are concentrated in economically and technologically developed areas such as North America, Europe, and China (Fink et al., 2020; Ma et al., 2020; Noton, 1998). Europe introduced the Water Framework Directive in late 2000, North America signed the Great Lakes Water Quality Agreement (GLWQA) in 1972 and launched the Clean Water Act in 2003 (Andreen, 2003; Kallis and Butler, 2001; Muldoon and Botts, 2005). These water quality management policies have effectively limited the further deterioration of lake water quality (Botts et al., 2018). However, our results showed that the algal blooms may increase significantly in lakes in China and the United States in the coming decades, which is inconsistent with the trend of surface water quality changes and long-term water quality management goals in these regions. In other words, as climate change intensifies and human activities increase, human efforts to control surface water quality may be offset by worsening algal blooms. Therefore, preventing algal blooms from getting out of control in lakes worldwide will become an important issue in future surface water quality management.

4.2. Advantages and limitations of the algorithm

We developed a machine-learning model based on random forests to project global algal blooms in large lakes. It is important to note that the machine learning model used in this study is a "black box" model as it did not explicitly demonstrate the mechanisms of interaction between

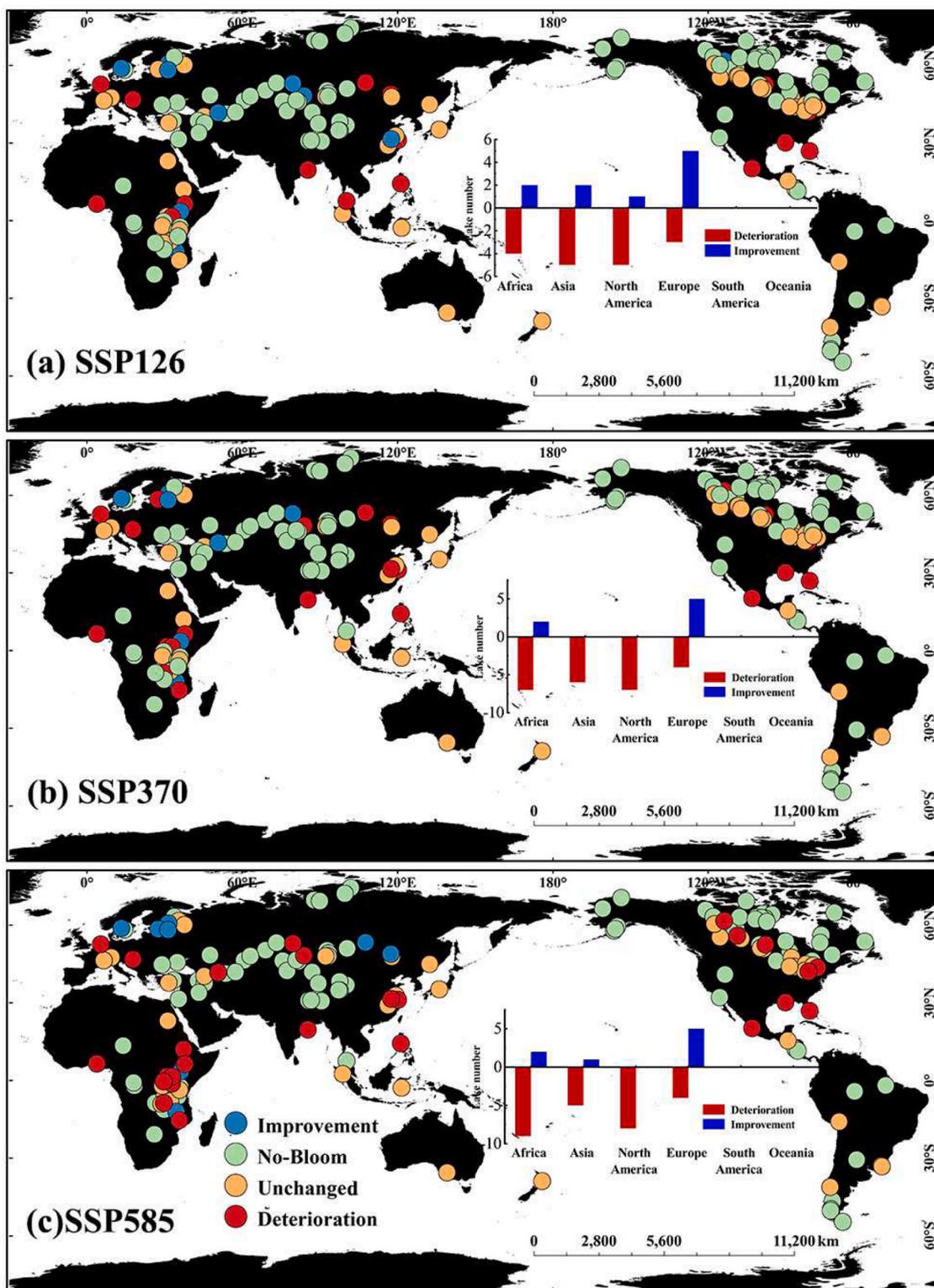


Fig. 7. The spatial distribution of the trends of average months with algal blooms in large global lakes under different SSP scenarios during 2020–2100. The bars in the figure show the trends of the lakes, with deteriorating (increasing in the average months with algal blooms) and restoring (decreasing in the average months with algal blooms).

parameters or the error propagation process (Iames et al., 2021; Kakouei et al., 2021; Piccolroaz et al., 2024). However, to improve the interpretability of the algorithm, this study incorporates factors such as meteorology, human activities, and lake morphology features based on the known mechanisms that drive algal blooms. For instance, the driving forces that directly affect algal growth (such as temperature and

nutrient-related factors) (Huisman et al., 2018) and the driving forces that have a significant influence on algal upwelling and the formation of algal scum (such as wind speed) (Zhang et al., 2006) were adopted in this study.

The Random Forest model was employed in this study, and a relative contribution test on the input variables (Fig. S2) was conducted to

understand how the model works and how each variable contributes to the projection model. The results revealed that the GDP and the population density had the highest contributions to the algorithm, reaching 14 % and 12 %, respectively. The land coverage rate of cropland within the basin contributed 6 %. The DSR related to lake morphology contributed 8 %, while lake area contributed 7 %. Among the meteorological factors, air pressure and accumulated temperature from the previous 1-2 months had the highest contributions, exceeding 5 %, while the contributions of other variables were relatively low. These results provide quantitative information on the relationship between relevant drivers and algal blooms for projection models. Variables representing human activity intensity had higher contribution rates, indicating that the mechanism of algal blooms caused by the input of nutrients and other substances resulting from human activities received more emphasis in the algorithm, thereby linking to the occurrence of algal blooms (Chen et al., 2010; Zhang et al., 2023). The contribution rates of lake morphology variables and meteorological variables were slightly lower. However, the results also reflected the sensitivity to external factors and the contribution of meteorological and environmental elements to algal blooms. This helped quantify the relationship between algal blooms and driving forces in complex nonlinear relationships.

In addition, the representativeness of the input data used in this study may be limited. The algorithm's input data was derived from the 161 lakes included in this study, so its application is confined to these lakes, and it remains unclear whether it can be generalized to other lakes. Future studies could enhance prediction models by incorporating data from a broader range of lakes across different regions and types. Besides, the range of possible input data values may also be limited. In future simulations with different SSPs and GCMs, some parameters may shortly exceed the input data's value range. Therefore, the relationship between these driving data and algal blooms is poorly established. Furthermore, the study lacks a clear explanation of the error propagation process and the specific roles of the parameters, making it difficult to understand the errors associated with the obtained results.

5. Conclusions

The present study developed and applied a novel method to project the future trends of algal blooms in large lakes across the globe from 2020 to 2100. The results showed that the number of lakes with algal blooms will increase globally under SSP370 and SSP585. The lakes with increasing algal blooms are mainly distributed in Africa, Asia, and North America, while lakes with decreasing occurrence are primarily found in Europe. Besides, the projected algal bloom trend is inconsistent with the reported trend of water quality in some regions, which indicates that the increased algal blooms could undermine efforts to control surface water quality. Therefore, preventing algal blooms will be more critical for surface water quality management in the future.

Data availability

HydroLAKES and HydroBASINS datasets were obtained from the HydroSHEDS website (<https://www.hydrosheds.org/>). The ERA5-Land dataset and MCD12Q1 dataset were acquired from the Google Earth Engine platform (<https://code.earthengine.google.com/>). The LandScan population density dataset was obtained from Oak Ridge National Laboratory (<https://landscan.ornl.gov/>). The ISIMIP3b simulation dataset was acquired from the Inter-Sectoral Impact Model Intercomparison Project (<https://www.isimip.org/>). The simulated land use/land cover dataset was obtained from the Northwest Pacific National Laboratory in the United States (<https://doi.org/10.25584/data.2020-07.1357/1644253/>). The simulated population density dataset was obtained from the School of Architecture at Tsinghua University (<https://doi.org/10.6084/m9.figshare.19608594.v2/>). The simulated gross domestic product (GDP) dataset was acquired from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy

of Sciences (<http://doi.org/10.5281/zenodo.4350027/>).

CRediT authorship contribution statement

Jinge Ma: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Hongtao Duan:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Cheng Chen:** Writing – review & editing, Software, Methodology, Data curation. **Zhigang Cao:** Writing – review & editing, Software, Methodology, Data curation. **Ming Shen:** Software, Resources, Methodology. **Tianci Qi:** Software, Resources, Methodology. **Qiuwen Chen:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Acknowledgments

The study is supported by the National Key Research and Development Program of China (2023YFC3208903), the National Nature Science Foundation of China (52121006, 52279071, 42101378), and China Yangtze Power Co., Ltd (Z242302052). Qiuwen Chen acknowledges the support from the XPLOER PRIZE. Cheng Chen acknowledges the support from the Cooperative Innovation Center for Water Safety and Hydro Science, Hohai University.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.watres.2024.122889](https://doi.org/10.1016/j.watres.2024.122889).

References

- Amer, M., Daim, T.U., Jetter, A., 2013. A review of scenario planning. *Futures*. 46, 23–40.
- Andren, W.L., 2003. Water quality today-has the clean water act been a success. *Ala. Law Rev.* 55, 537.
- Bonshoms, M., Ubeda, J., Liguori, G., Körner, P., Navarro, Á., Cruz, R., 2022. Validation of ERA5-Land temperature and relative humidity on four Peruvian glaciers using on-glacier observations. *J. Mt. Sci.* 19 (7), 1849–1873.
- Botts, L., Muldoon, P., Botts, P., von Moltke, K., 2018. Knowledge, Power, and Participation in Environmental Policy Analysis. Routledge, pp. 121–144.
- Burford, M.A., Carey, C.C., Hamilton, D.P., Huisman, J., Paerl, H.W., Wood, S.A., Wulff, A., 2020. Perspective: advancing the research agenda for improving understanding of cyanobacteria in a future of global change. *Harmful Algae* 91, 101601.
- Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sens. Environ.* 248, 111974.
- Carstensen, J., Henriksen, P., Heiskanen, A.S., 2007. Summer algal blooms in shallow estuaries: definition, mechanisms, and link to eutrophication. *Limnol. Oceanogr.* 52 (1), 370–384.
- Chen, C., Chen, Q., Yao, S., He, M., Zhang, J., Li, G., Lin, Y., 2024. Combining physical-based model and machine learning to forecast chlorophyll-a concentration in freshwater lakes. *Sci. Total Environ.* 907, 168097.
- Chen, M., Vernon, C.R., Graham, N.T., Hejazi, M., Huang, M., Cheng, Y., Calvin, K., 2020. Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. *Sci. Data* 7 (1), 320.
- Chen, N., Zhang, Y., Li, Y., 2010. An integrated analysis of dynamic characteristics of harmful algal bloom in fresh water in China. *Ecol. Environ. Sci.* 19 (8), 1994–1998.
- Chen, X., Jiang, H., Sun, X., Zhu, Y., Yang, L., 2016. Nitrification and denitrification by algae-attached and free-living microorganisms during a cyanobacterial bloom in Lake Taihu, a shallow Eutrophic Lake in China. *Biogeochemistry* 131 (1), 135–146.
- Couture, R.M., Tominaga, K., Starrfelt, J., Moe, S.J., Kaste, O., Wright, R.F., 2014. Modelling phosphorus loading and algal blooms in a Nordic agricultural catchment-lake system under changing land-use and climate. *Environ. Sci. Process. Impacts* 16 (7), 1588–1599.
- Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. *Photogramm. Eng. Remote Sens.* 66 (7), 849–857.

- Elliott, J.A., 2012. Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. *Water Res.* 46 (5), 1364–1371.
- Fang, C., Song, K., Paerl, H.W., Jacinthe, P.A., Wen, Z., Liu, G., Tao, H., Xu, X., Kutser, T., Wang, Z., Duan, H., Shi, K., Shang, Y., Lyu, L., Li, S., Yang, Q., Lyu, D., Mao, D., Zhang, B., Cheng, S., Lyu, Y., 2022. Global divergent trends of algal blooms detected by satellite during 1982–2018. *Glob. Change Biol.* 28, 2327–2340.
- Fink, G., Burke, S., Simis, S.G.H., Kangur, K., Kutser, T., Mulligan, M., 2020. Management options to improve water quality in lake peipis: insights from large scale models and remote sensing. *Water Resour. Manag.* 34 (7), 2241–2254.
- Guan, W., Bao, M., Lou, X., Zhou, Z., Yin, K., 2022. Monitoring, modeling and projection of harmful algal blooms in China. *Harmful Algae* 111, 102164.
- Ho, J., Michalak, A., Pahlevan, N., 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* 574, 667–670.
- Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., Zheng, C., 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. *Nat. Geosci.* 15, 130–134.
- Huang, C., Li, Y., Yang, H., Sun, D., Yu, Z., Zhang, Z., Chen, X., Xu, L., 2014. Detection of algal bloom and factors influencing its formation in Taihu Lake from 2000 to 2011 by MODIS. *Environ. Earth Sci.* 71 (8), 3705–3714.
- Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* 16 (8), 471–483.
- Iames, J.S., Salls, W.B., Mehaffey, M.H., Nash, M.S., Christensen, J.R., Schaeffer, B.A., 2021. Modeling anthropogenic and environmental influences on freshwater harmful algal bloom development detected by MERIS over the central United States. *Water Resour. Res.* 57 (10), e2020WR028946.
- Janssen, A.B.G., Janse, J.H., Beusen, A.H.W., Chang, M., Harrison, J.A., Huttunen, I., Kong, X., Rost, J., Teurlinck, S., Troost, T.A., van Wijk, D., Mooij, W.M., 2019. How to model algal blooms in any lake on earth. *Curr. Opin. Environ. Sustain.* 36, 1–10.
- Jansson, L.H.A.M., 1985. Principles of lake sedimentology. *Int. Rev. Gesamten Hydrobiol. Hydrogr.* 70 (3), 431–431.
- Jochimsen, E.M., Carmichael, W.W., An, J.S., Cardo, D.M., Cookson, S.T., Holmes, C.E. M., Antunes, M.B.D., de Melo, D.A., Lyra, T.M., Barreto, V.S.T., Azevedo, S., Jarvis, W.R., 1998. Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *N. Engl. J. Med.* 338 (13), 873–878.
- Jones, E.R., Bierkens, M.F.P., van Puijenbroek, P.J.T.M., van Beek, L.P.H., Wanders, N., Sutanudjaja, E.H., van Vliet, M.T.H., 2023. Sub-Saharan Africa will increasingly become the dominant hot spot of surface water pollution. *Nat. Water* 1 (7), 602–613.
- Kakouei, K., Kraemer, B.M., Anneville, O., Carvalho, L., Feuchtmayr, H., Graham, J.L., Higgins, S., Pomati, F., Rudstam, L.G., Stockwell, J.D., Thackeray, S.J., Vanni, M.J., Adrian, R., 2021. Phytoplankton and cyanobacteria abundances in mid-21st century lakes depend strongly on future land use and climate projections. *Glob. Change Biol.* 27 (24), 6409–6422.
- Kallis, G., Butler, D., 2001. The EU water framework directive: measures and implications. *Water Policy* 3 (2), 125–142.
- Kutser, T., 2004. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. *Limnol. Oceanogr.* 49 (6), 2179–2189.
- Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Process.* 27 (15), 2171–2186.
- Li, X., Messina, J.P., Moore, N.J., Fan, P., Shortridge, A.M., 2017. MODIS land cover uncertainty in regional climate simulations. *Clim. Dyn.* 49 (11), 4047–4059.
- Lin, S.Q., Pierson, D.C., Mesman, J.P., 2023. Prediction of algal blooms via data-driven machine learning models: an evaluation using data from a well-monitored mesotrophic lake. *Geosci. Model. Dev.* 16 (1), 35–46.
- Ma, J., Loiselle, S., Cao, Z., Qi, T., Shen, M., Luo, J., Song, K., Duan, H., 2023. Unbalanced impacts of nature and nurture factors on the phenology, area and intensity of algal blooms in global large lakes: MODIS observations. *Sci. Total Environ.* 880, 163376.
- Ma, T., Zhao, N., Ni, Y., Yi, J., Wilson, J.P., He, L., Du, Y., Pei, T., Zhou, C., Song, C., Cheng, W., 2020. China's improving inland surface water quality since 2003. *Sci. Adv.* 6 (1).
- McCrackin, M.L., Jones, H.P., Jones, P.C., Moreno-Mateos, D., 2017. Recovery of lakes and coastal marine ecosystems from eutrophication: a global meta-analysis. *Limnol. Oceanogr.* 62 (2), 507–518.
- Merder, J., Harris, T., Zhao, G., Stasinopoulos, D.M., Rigby, R.A., Michalak, A.M., 2023. Geographic redistribution of microcystin hotspots in response to climate warming. *Nat. Water* 1 (10), 844–854.
- Muldoon, P., Botts, L., 2005. Evolution of the Great Lakes Water Quality Agreement. Michigan State University Press.
- Mullin, C.A., Kirchhoff, C.J., Wang, G., Vlahos, P., 2020. Future projections of water temperature and thermal stratification in connecticut reservoirs and possible implications for cyanobacteria. *Water Resour. Res.* 56 (11), e2020WR027185.
- Muñoz Sabater, J., 2019. ERA5-Land Monthly Averaged Data from 1981 to Present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi:10.24381/cds.68d2bb3.
- Noton, L., 1998. Water quality of Lesser Slave Lake and its tributaries, 1991–93, AlbertaEnvironment, Edmonton, Alberta 88p.
- OrihelDiane, M., BirdDavid, F., BrylinskyMichael, C.H., DonaldDerek, B., HuangDorothy, Y., GianiAlessandra, K.D., KlingHedy, K.G., LeavittPeter, R., NielsenCharlene, C., ReedykSharon, R.R.C., WatsonSue, B., ZurawellRon, W., VinebrookeRolf, D., SmithRalph, E.H., 2012. High microcystin concentrations occur only at low nitrogen-to-phosphorus ratios in nutrient-rich Canadian lakes. *Can. J. Fish. Aquat. Sci.* 69 (9), 1457–1462.
- Paerl, H.W., Gardner, W.S., Havens, K.E., Joyner, A.R., McCarthy, M.J., Newell, S.E., Qin, B., Scott, J.T., 2016. Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. *Harmful Algae* 54, 213–222.
- Paerl, H.W., Huisman, J., 2008. Blooms like it hot. *Science* 320 (5872), 57–58 (1979).
- Pelletier, J.D., Brad Murray, A., Pierce, J.L., Bierman, P.R., Breshears, D.D., Crosby, B.T., Ellis, M., Foufloula-Georgiou, E., Heimsath, A.M., Houser, C., Lancaster, N., Marani, M., Merritts, D.J., Moore, L.J., Pederson, J.L., Poulos, M.J., Rittenour, T.M., Rowland, J.C., Ruggiero, P., Ward, D.J., Wickert, A.D., Yager, E.M., 2015. Forecasting the response of Earth's surface to future climatic and land use changes: a review of methods and research needs. *Earths Future* 3 (7), 220–251.
- Piccolroaz, S., Zhu, S., Ladwig, R., Carrea, L., Oliver, S., Piotrowski, A.P., Ptak, M., Shinohara, R., Sojka, M., Woolway, R.I., Zhu, D.Z., 2024. Lake water temperature modeling in an era of climate change: data sources, models, and future prospects. *Rev. Geophys.* 62 (1), e2023RG000816.
- Qin, B., Zhu, G., Gao, G., Zhang, Y., Li, W., Paerl, H.W., Carmichael, W.W., 2010. A drinking water crisis in Lake Taihu, China: linkage to climatic variability and lake management. *Environ. Manag.* 45 (1), 105–112.
- Ralston, D.K., Moore, S.K., 2020. Modeling harmful algal blooms in a changing climate. *Harmful Algae* 91, 101729.
- Reichwaldt, E.S., Ghadouani, A., 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in the changing climate: between simplistic scenarios and complex dynamics. *Water Res.* 46 (5), 1372–1393.
- Shen, M., Cao, Z., Xie, L., Zhao, Y., Qi, T., Song, K., Lyu, L., Wang, D., Ma, J., Duan, H., 2023. Microcystins risk assessment in lakes from space: implications for SDG 6.1 evaluation. *Water Res.* 245, 120648.
- Shen, M., Cao, Z., Xue, K., Liu, D., Qi, T., Ma, J., Duan, H., 2022a. Natural and human activities driving the spatiotemporal variability of water clarity in lakes across Eastern China. *Int. J. Appl. Earth Observ. Geoinform.* 114, 103037.
- Shen, M., Luo, J., Cao, Z., Xue, K., Qi, T., Ma, J., Liu, D., Song, K., Feng, L., Duan, H., 2022b. Random forest: an optimal chlorophyll-a algorithm for optically complex inland water suffering atmospheric correction uncertainties. *J. Hydrol.* 128685 (Amst).
- Sulla-Menashe, D. and Friedl, M., 2019. MCD12Q1 MODIS/Terra+ aqua land cover type yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. doi:10.5067/MODIS/MCD12C1.006.
- Tellman, B., Sullivan, J.A., Kuhn, C., Kettner, A.J., Doyle, C.S., Brakenridge, G.R., Erickson, T.A., Slayback, D.A., 2021. Satellite imaging reveals increased proportion of population exposed to floods. *Nature* 596 (7870), 80–86.
- Tewari, M., Kishitawal, C.M., Moriarty, V.W., Ray, P., Singh, T., Zhang, L., Treinish, L., Tewari, K., 2022. Improved seasonal prediction of harmful algal blooms in Lake Erie using large-scale climate indices. *Commun. Earth. Environ.* 3 (1), 195.
- Tong, Y., Feng, L., Wang, X., Pi, X., Xu, W., Woolway, R.I., 2023. Global lakes are warming slower than surface air temperature due to accelerated evaporation. *Nat. Water* 1 (11), 929–940.
- Wang, H., Xu, C., Liu, Y., Jeppesen, E., Svenning, J.C., Wu, J., Zhang, W., Zhou, T., Wang, P., Nangombe, S., 2021. From unusual suspect to serial killer: cyanotoxins boosted by climate change may jeopardize African megafauna. *Innovation*, 100092.
- Wang, J., Liu, X., Beusen, A.H.W., Middelburg, J.J., 2023. Surface-Water Nitrate Exposure to World Populations Has Expanded and Intensified during 1970–2010. *Environ. Sci. Technol.* 57 (48), 19395–19406.
- Wang, T., Sun, F., 2021. Spatially explicit global gross domestic product (GDP) data set consistent with the Shared Socioeconomic Pathways. *Earth Syst. Sci. Data Discuss.* 2021, 1–34.
- Wang, X., Meng, X., Long, Y., 2022. Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. *Sci. Data* 9 (1), 563.
- Wilkinson, G.M., Walter, J.A., Buelo, C.D., Pace, M.L., 2021. No evidence of widespread algal bloom intensification in hundreds of lakes. *Front. Ecol. Environ.* 20 (1), 16–21.
- Woolway, R.I., Huang, L., Sharma, S., Lee, S.S., Rodgers, K.B., Timmermann, A., 2022. Lake ice will be less safe for recreation and transportation under future warming. *Earths. Future* 10 (10), e2022EF002907.
- Zang, N., Cao, Y., Wang, Y., Huang, B., Zhang, L., Mathiopoulou, P.T., 2021. Land-use mapping for high-spatial resolution remote sensing image via deep learning: a review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 5372–5391.
- Zhang, C., Kong, X., Xue, B., Zhao, C., McGowan, S., Lin, Q., Zhang, K., Shen, J., 2023. Double-edged effects of anthropogenic activities on lake ecological dynamics in northern China: evidence from palaeolimnology and ecosystem modelling. *Freshw. Biol.* 68 (6), 940–955.
- Zhang, Y., Qin, B., Zhu, G., Gao, G., Luo, L., Chen, W., 2006. Effect of sediment resuspension on underwater light field in shallow lakes in the middle and lower reaches of the Yangtze River: a case study in Longgan Lake and Taihu Lake. *Sci. China Ser. D Earth Sci.* 49, 114–125.
- Zhao, X., Liu, Y., Guo, Y.M., Xu, C., Chen, L., Codd, G.A., Chen, J., Wang, Y., Wang, P.Z., Yang, L.W., Zhou, L., Li, Y., Xiao, S.M., Wang, H.J., Paerl, H.W., Jeppesen, E., Xie, P., 2023. Meta-analysis reveals cyanotoxins risk across African inland waters. *J. Hazard. Mater.* 451, 131160.
- Zhou, Z.X., Yu, R.C., Zhou, M.J., 2022. Evolution of harmful algal blooms in the East China Sea under eutrophication and warming scenarios. *Water Res.* 221, 118807.