Geographic Analysis of the Vulnerability of U.S. Lakes to Cyanobacterial Blooms under Future Climate

JONATHAN B. BUTCHER, a MARK FERNANDEZ, a THOMAS E. JOHNSON, b AFSHIN SHABANI, a AND SYLVIA S. LEE b

a Tetra Tech, Inc., Research Triangle Park, North Carolina
b Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

(Manuscript received 16 March 2023, in final form 2 November 2023, accepted 6 November 2023)

ABSTRACT: Cyanobacteria blooms are an increasing concern in U.S. freshwaters. Such blooms can produce nuisance conditions, deplete oxygen, and alter the food chain, and in some cases they may produce potent toxins, although many factors may modulate the relationships between biomass and toxin production. Cyanobacterial blooms are in turn associated with nutrient enrichment and warm water temperatures. Climate change is expected to increase water temperatures and, in many areas, surface runoff that can transport nutrient loads to lakes. While some progress has been made in short-term prediction of cyanobacterial bloom and toxin risk, the long-term projections of which lakes will become more vulnerable to such events as a result of climate change is less clear because of the complex interaction of multiple factors that affect bloom probability. We address this question by reviewing the literature to identify risk factors that increase lake vulnerability to cyanobacterial blooms and evaluating how climate change may alter these factors across the sample of conterminous U.S. lakes contained in the 2007 National Lakes Assessment. Results provide a national-scale assessment of where and in which types of lakes climate change will likely increase the overall risk of cyanobacterial blooms, rather than finer-scale prediction of expected cyanobacterial and toxin levels in individual lakes. This information can be used to guide climate change adaptation planning, including monitoring and management efforts to minimize the effects of increased cyanobacterial prevalence.

SIGNIFICANCE STATEMENT: Cyanobacteria blooms and associated algal toxins are an increasing problem in U.S. freshwater lakes and reservoirs. Climate change may further increase bloom frequency and severity. We survey the literature on relationships between bloom formation and climate. These relationships are combined with projections of future climate and lake response to develop indices of where and in what types of lakes such blooms are most likely to increase relative to current conditions. The results can help to focus monitoring and management measures to mitigate potential impacts on human health, wildlife, and aquatic biota.

KEYWORDS: Inland seas/lakes; Climate prediction; Ecological models; Adaptation; Ecosystem effects; Lake effects

1. Introduction

Harmful algal blooms (HABs), primarily associated with cyanobacteria, are a problem of increasing concern for water quality, domestic animals, and human health throughout the United States. Toxic substances produced by some cyanobacteria (cyanotoxins), are of particular concern, but the ability to forecast toxin production, especially over longer temporal scales, is still limited. The occurrence of cyanobacterial blooms, however, is used to alert lake managers of the increased risk of cyanotoxins (Chorus and Welker 2021).

In this paper we discuss the vulnerability of freshwater lakes and reservoirs to cyanobacterial blooms. There is no generally accepted quantitative definition of algal blooms, so we use the imprecise empirical definition where bloom refers to the occurrence of excess algal growth sufficient to threaten support of one or more beneficial uses of a waterbody. A cyanobacterial bloom is defined ad hoc as a bloom in which more than 50% of the algal biovolume is cyanobacteria (after Downing et al. 2001). Vulnerability is the propensity of a waterbody to suffer adverse effects relative to a future stressor (Cardona et al. 2012), where the adverse effects in the context of cyanobacterial blooms may range from filter-clogging scums to production of algal toxins. In contrast, risk is the product of the stressor distribution and its consequences (Birkmann 2007). We focus on vulnerability to climate change. Defining vulnerability qualitatively is sufficient because the analysis examines relative changes in susceptibility not quantitative estimates of risk. We refer to stressors and waterbody characteristics that are associated with increasing vulnerability as risk factors, using the term factors rather than hypotheses as the relationships may reflect correlation rather than causation. Last, we define risk metrics as summary indices that are used to score the future change in the relative vulnerability of lakes to cyanobacterial blooms. Increased nutrient concentrations, warmer water temperatures, and a stable water column with low mixing and flushing rates are known factors associated with an increased likelihood

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/El-D-23-0004.s1.

Corresponding author: Jonathan B. Butcher, jon.butcher@tetratech.com

Earth Interactions is published jointly by the American Meteorological Society, the American Geophysical Union, and the Association of American Geographers.

DOI: 10.1175/El-D-23-0004.1 e230004

© 2023 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
of cyanobacterial blooms in lakes (Paerl 1988; Rousso et al. 2020; Paerl and Huisman 2009; Huisman et al. 2018; O’Neil et al. 2012). Throughout the United States, climate change over the next century is expected to result in increased average water temperatures, alterations to flushing and water column mixing patterns, and changes in the timing and amount of the delivery of nutrient loads to inland lakes (U.S. Global Change Research Program (USGCRP) 2018). If realized, these changes will increase the vulnerability of inland lakes to cyanobacterial blooms (although not necessarily correct for all lakes).

Increases in cyanobacterial blooms have already been observed and attributed to climate-induced effects. A recent study (Smucker et al. 2021) documents increasing bloom frequency through analysis of 30 years of data on 20 U.S. reservoirs, concluding that cyanobacteria cell densities, including those capable of producing toxins, have rapidly increased since 2005 with blooms quickly becoming recurrent problems within the span of a decade. Similar results are reported for Canada (Winter et al. 2011). The researchers found no clear or consistent changes in land use, precipitation, and nutrient concentrations among these reservoirs, but did find that reservoirs are heating up earlier in the year and remaining warm longer than they did 30 years ago. While blooms increased in all study reservoirs (20), the greatest increases were in reservoirs with urban or agriculturally dominated watersheds, suggesting that the effects of warming and changes in seasonality become stronger in reservoirs with more nutrient runoff. Satellite-derived estimates also show a consistent increase in the peak intensity of blooms in lakes worldwide since the 1980s (e.g., Ho et al. 2019).

Over the last decade there has been an exponential increase in the scientific literature related to cyanobacterial bloom vulnerability and prediction in lakes. These studies can be divided into two general classes: process-based analyses that attempt to predict cyanobacterial response based on causal models, and data-driven analyses, based on correlation of blooms to observed environmental factors (Rousso et al. 2020). The process-based analyses have been primarily used to evaluate longer term, future scenarios while the data-driven approaches have tended to focus on short-term (days to weeks) forecasts of bloom occurrence that can provide early warning to water utilities and other users of lakes.

The likely effects of climate change on cyanobacterial blooms have been addressed both in a general, theoretical context, and in case study scenario analysis of individual lakes. A detailed systematic literature review was undertaken by Rousso et al. (2020) with additional reviews and analyses by Wang et al. (2023). Other more geographically limited studies of cyanobacterial bloom vulnerability across multiple lakes have been reported for Georgia (United States) (Weber et al. 2020), Europe (Elliott 2012; Mantzouki et al. 2018), and Argentina (O’Farrell et al. 2019).

While many authors have identified similar risk factors related to cyanobacterial blooms, the inferences are not consistent, reflecting the complex interaction of factors that lead to dominance of cyanobacteria over other algal groups. For example, elevated water temperatures favored cyanobacterial blooms in some systems but not in others. High nutrient levels [primarily phosphorus (P)] favor phytoplankton blooms in general, but their effect on cyanobacteria populations seems to depend on the timing of nutrient load delivery relative to ice out, stratification onset, and other factors. Onset of a stable, stratified water column may give an advantage to buoyancy regulating cyanobacteria relative to other phytoplankton but may also suppress blooms depending on the depth of the mixed layer and its impact on summer nutrient availability (Huisman et al. 2018; Paerl and Paul 2012; O’Neil et al. 2012). In addition to site-specific physical and chemical factors, it is also important to note that cyanobacteria consist of a vast number of taxa with a range of characteristics that affect interspecies competition, for instance, giving an advantage to cyanobacteria that can fix atmospheric nitrogen (N) in systems where dissolved N becomes limiting for algal growth (Rigosi et al. 2014).

Chapra et al. (2017) undertook a screening-level, process-based modeling study of HABs in 310 large reservoirs and lakes throughout the contiguous United States using linked models of future climate, rainfall–runoff, water demand, reservoir systems behavior, and cyanobacterial response. They concluded that the mean number of cyanobacteria HAB events would increase on average from 7 days per year per waterbody to 18–39 days per year by the end of the century. However, the results are dependent on assumptions built into the model as to how cyanobacteria will respond to temperature, nutrient load, vertical stratification, pH, and other factors, and these inherent assumptions may bias the results. More generally, site-specific and cyanobacteria taxon-specific complexities and interactions make it difficult to draw general conclusions about the effects of climate change on cyanobacterial bloom vulnerability from process-based analyses.

In our study, we focus on the long-term increase in vulnerability of freshwater lakes and reservoirs in the contiguous United States to increased cyanobacterial blooms associated with projected climate change. The emphasis is on changes in vulnerability, not quantitative prediction of bloom probability, with the intention of identifying those lakes most prone to increased future blooms where mitigation and adaptation responses may be needed. The current state of the science does not provide a reliable basis for predicting long-term, climate-induced changes in bloom vulnerability on an individual lake basis. However, many of the factors that increase vulnerability are known, although full, quantitative understanding of how they interact may be lacking. Accordingly, our methodology is based on a qualitative meta-analysis to first identify the key risk factors from the literature for increased bloom vulnerability, and then provides a spatially explicit evaluation of which geographic areas and what lake characteristics have the greatest vulnerability for increased cyanobacterial blooms under future climate based on those factors.

2. Methods

Our approach involves two general steps:

1) identify and evaluate the evidence supporting proposed risk factors that suggest how cyanobacteria bloom vulnerability
can, either directly or indirectly, be affected by changes in climate, and
2) develop methods to predict future changes in key physical characteristics of individual lakes (e.g., water temperature and stratification) in response to projected climate change.

The outcome of these analyses is then presented in section 4.

a. Identifying risk factors

The first step in the methodology was to identify and evaluate the evidence supporting different risk factors related to lake and reservoir vulnerability to cyanobacterial blooms. Two recent reviews of the HABs/climate topic suggest that quantitative meta-analysis may not be feasible at this time due to the complexity of the interacting factors that lead to cyanobacterial blooms and the current state of understanding of the issue (Rousso et al. 2020; Wang et al. 2023). We therefore do not attempt a quantitative meta-analysis.

As suggested in the introduction section, the relationship between cyanobacterial bloom vulnerability and physical and chemical drivers is highly complex and variable, and process-based tools that can adequately predict such vulnerability across multiple lake types have met with limited success. Further, most papers on the subject are based on studies of one or a limited number of lakes or are based on projections by simulation models with a low level of precision. Few controlled experiments on bloom formation exist [although see Salk et al. (2022) and Molot et al. (2021)]. This sort of information is not amenable to classical meta-analysis, which tests hypotheses based on the statistical combination of information on a specific target of interest from multiple quantitative experiments with different effect sizes. This is a problem that has often been faced in ecological studies and other fields that attempt to combine observation-based data rather than experimental quantitative data with a testable hypothesis. The problem is particularly acute when attempting to make inferences about the future with changed conditions under which past assumptions about stable reference conditions may no longer be relevant. For such situations it is important to focus on uncertainty or confidence relevant to management decisions based on qualitative evaluations of potential responses (West et al. 2012).

The available literature is more amenable to qualitative meta-analysis or metasynthesis. For the qualitative meta-analysis, we started with the relevant literature identified in the systematic review of Rousso et al. (2020) and the follow-on study of Wang et al. (2023) and filtered them to lentic freshwater systems. We conducted additional literature searches to identify more recently published articles relevant to the topic in peer-reviewed journals as well as some sources not included by Rousso or Wang. This resulted in the selection of 96 studies in 97 articles, which are listed in the online supplemental material.

The 96 studies were categorized by primary type of information class (Salafsky et al. 2019), as summarized in Table 1. Articles were read in their entirety twice and information relevant to vulnerability to cyanobacterial blooms was extracted in the form of risk factors that relate an outcome to an exposure in a defined population. For example, “In deeper lakes that stratify [population] increased air temperatures [stressor] is associated with earlier and stronger vertical stratification that enhances risk of blooms of buoyancy-regulating cyanobacteria [outcome].” In summarizing such statements care was used to identify secondary factors that explain or infer how the relationship is affected by other modifying factors (Bennett et al. 2017). For example, the suggested relationships between air temperature, enhanced stratification, and cyanobacteria response may not hold if counteracting factors such as increased wind or higher summer inflows are present.

b. Predicting future changes in physical characteristics of individual lakes

The second step involves predicting future changes in key, physical characteristics of individual lakes (e.g., water temperature, stratification strength, frequency of mixing, average date of onset of stratification, and other characteristics that we refer to as indicators) in response to projected climate change. Evaluating changes on a spatial basis makes use of prior work on estimated responses of lake water temperature and vertical stability to climate forcing as a function of lake characteristics using the Lake Ice Snow and Sediment Simulator (LISSS) model (Subin et al. 2012) as described in Butcher et al. (2015, 2017). The LISSS model simulations in Butcher et al. (2015) were conducted using site-specific hourly climate data derived from regional climate model (RCM) applications forced by global climate model (GCM) simulations of future climate conducted by the North American Regional Climate Change Assessment Program (NARCCAP). Such data are available only for a small subset of GCMs, while commonly used statistical downscaling products do not provide subdaily output and do not include various weather time series data required by LISSS, making direct application to multiple lakes a difficult task. Butcher et al. (2017) therefore developed a surrogate modeling approach using artificial intelligence (AI) tools to predict the results of the LISSS simulations based on monthly climate data.

To fit surrogate models of the indicators that can be applied to future climate we used the sensitivity simulations conducted for multiple climate scenarios and a continuous range of lake depth, surface area, and water transparency using a Sobol sampling scheme as reported in Butcher et al. (2017).
Sobol sampling uses a quasi-random space-filling sequence proposed by (Sobol' 1967) and refined by (Antonov and Saleev 1979) to populate the test data. In a Sobol sequence, data gaps are filled with successive sample realizations, maximizing coverage of the response surface sample space. The results of those simulation runs were used in the current work to estimate indicators for each lake as a function of current and future monthly climate variables using a machine learning surrogate of the original process-based model results.

While Butcher et al. (2017) used Gaussian processes AI for this purpose in predicting summer water temperatures at the thermocline they also found that comparable results could be obtained through use of random forest regression. Random forest (RF) is an ensemble learning technique that uses a bootstrapping approach to generate multiple decision trees that relate outcomes to predictors. While each individual tree has high uncertainty, aggregating the results of many individual trees reduces variance and can provide a strong empirical basis for prediction.

We evaluated direct application of RF to outcomes of interest and found that predictions could be improved through use of the related method of boosted regression trees (BRT; Elith et al. 2008; Chen and Guestrin 2016), which combine gradient boosting with the same type of decision trees used in RF regression. Boosted regression trees combine the strengths of regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance). Unlike RF, the decision trees in a gradient boosting are built additively, with each new tree designed to improve on the deficiencies of the previous trees. The trees are aggregated as they are processed, typically resulting in stronger predictive performance than RF, which combines trees at the end. Another advantage is that the BRT approach typically does not require transformation of the variables to be predicted. We used the R (R Core Team 2021) software packages ranger (Wright and Ziegler 2017) and xgboost (Chen et al. 2022) to implement RF regression and BRT.

Each BRT model was tuned using fivefold cross validation. Learning rate, maximum tree depth, minimum leaf weight, gamma, and lambda hyperparameters were tuned using an iterative grid search method in which the average holdout root-mean-square error (RMSE) was used as the criterion and gamma and lambda are regularization parameters to help prevent overfitting. The number of trees was determined using stopping rules related to holdout data performance. Model fit statistics are calculated from test data, which consist of a 30% random sample from the total dataset not used in model fitting. We stratified the test data on geographic region and climate scenario to ensure the test dataset is equally representative across these variables.

3. Data

a. Lake database

For a representative sample of lakes and their morphometric characteristics we use the 2007 National Lakes Assessment (NLA) database (U.S. Environmental Protection Agency 2009). The NLA is part of EPA’s National Aquatic Resource Surveys program, which aims to provide statistically valid data describing water resource quality conditions across the country. We use only the lakes included in the 2007 NLA to ensure consistency in reported variables. The database was filtered to include only lakes with depths from 2 to 30 m and surface area from 0.1 to 100 km² for which the simulation approach used is most appropriate (Butcher et al. 2015) and includes both lakes that stratify and ones that do not. This left a total of 961 lakes and reservoirs for analysis (Fig. 1).

b. Climate scenarios

We analyze vulnerability of the NLA lakes for historical conditions (1971–2000), future midcentury conditions (2041–70), and future end-century conditions (2070–99). Scenarios are based on modeling results from phase 5 of the Coupled Model Intercomparison Project (CMIP5) associated with the fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC). We use the multivariate adaptive constructed analogs (MACA; Abatzoglou and Brown 2012) statistically downscaled data for both historical and future conditions for a consistent evaluation of likely changes. The MACA dataset was chosen because, unlike other statistically downscaled climate products, it provides simultaneous downscaling of precipitation, temperature, humidity, wind, and radiation (rather than just precipitation and temperature), ensuring physical consistency between components of the energy balance. The spatial downscaling is at a resolution of (1/24)°.

As in Butcher et al. (2017) we use MACA v2-METDATA downscaled output of six GCMs under the representative concentration pathway 8.5 W m⁻² (RCP8.5) scenario that 1) are evaluated as having good skill on prediction of warm–dry conditions (i.e., conditions under which maximum lake thermal increase is expected) in temperate latitudes (Hao et al. 2013), 2) have high spatial resolution (Toreti et al. 2013), and 3) exhibit a range of projected increases in annual mean maximum air temperatures across the conterminous United States (Table 2). RCP8.5 was selected as a plausible worst-case scenario that could arise without additional efforts to constrain emissions (Pachauri et al. 2014).

While the climate datasets used are the same as in Butcher et al. (2017), we re-extracted data (at the centroid of each lake) from the MACA Themetic Real-time Environmental Distributed Data Services (THREDDS) server (http://maca.northwestknowledge.net/) to ensure that any corrections to the MACA archive were captured and to provide for retrieval of additional climate variables.

c. Other inputs

Land and reservoir vulnerability to increased cyanobacterial blooms is affected by existing nutrient supply while potential changes in nutrient loading rates are risk factors. To assess nutrient supply, we use results computed for each 2007 NLA lake watershed for N and P surplus generated for the EPA national nutrient inventories (Sabo et al. 2021b, 2019, 2021a);
results for individual NLA lakes were provided by R. D. Sabo (2022, personal communication).

Changes in nutrient loads are likely to be associated with changes in runoff. The MACA product provides information on future precipitation, but not future runoff. For an approximate indication of future changes in runoff, which is likely to drive P loading, we use climate adjusted monthly calculations from the variable infiltration capacity (VIC) model linked to the U.S. Bureau of Reclamation’s Bias Corrected and Statistically Downscaled climate product (BCSD: Archfield et al. 2020; Brekke et al. 2014) for the same six climate scenarios extracted from MACA. These were retrieved via ftp from the Lawrence Livermore National Laboratory Green Data Oasis (https://gdo-dcp.ucclnl.org). (VIC models have been generated to match the MACA CMIP5 gridded climate products; however, at this time publicly available websites serve this product only for the Pacific Northwest.)

We use the BCSD projections for surface (overland) runoff, rather than total water yield, as surface runoff is more closely associated with P loading and large flow events that may cause water column mixing. VIC is a gridded product and is available as both unrouted (local) and routed flows. However, the routing on a coarse grid is suspect, especially with regard to smaller lakes. We therefore use the local, unrouted runoff depth estimates for the grid cell coincident with the centroid of each NLA lake as an approximate indicator. That BCSD is at a coarser spatial scale than localized constructed analogs [LOCA; \( \frac{1}{64} \) vs \( \frac{1}{16} \)] is actually an advantage in regard to estimating approximate conditions surrounding an individual lake.

### 4. Results

#### a. Risk factors

Risk factors extracted from the literature are summarized in full in files S1 and S2 in the online supplemental material. Counts of proposed risk factors, summarized in Table 3, are dominated by nutrients (N and P) and water temperature, with most researchers finding that these factors promoted cyanobacterial blooms. The counts, however, should be evaluated in a qualitative sense only, as many researchers did not consider factors other than nutrients and temperature, and sometimes only one of those. Further, few studies report factors that did not have an effect and many of the studies cite one another to support their conclusions. The proposed risk factors are rated as to *evidence* (the amount of data presented to assess the strength of a relationship) and *agreement* (the level of consensus in the scientific community regarding the

| TABLE 2. MACAv2-METDATA downscaled CMIP5 GCMs used in this study. |
|---------------------|-------------------|
| **Model**            | **Institution**                             |
| BCC_CSM1.1           | Beijing Climate Center, China Meteorological Administration, China |
| CNRM-CM5             | Centre National de Recherches, Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France |
| GFDL-ESM2M           | NOAA/Geophysical Fluid Dynamics Laboratory, United States |
| HadGEM2-CC365        | Met Office Hadley Centre, United Kingdom |
| IPSL-CM5A-MR         | L’Institut Pierre-Simon Laplace, France |
| MRI-CGCM3            | Meteorological Research Institute, Japan |

FIG. 1. Location of lakes analyzed from 2007 NLA database.
Table 3. Summary of risk factors and weight of evidence identified in the literature for physical characteristics affecting lake vulnerability to cyanobacterial blooms. Note that stressor categories with only one entry are not included. A plus sign indicates that the response is associated with an increase in the stressor class; a minus sign indicates association with a decrease. COD is chemical oxygen demand. Evidence is rated high (>20 studies; H), medium (6–20 studies; M), or low (<6 studies; L). Agreement is rated as high (more than two-thirds of studies report the same effect direction; H) or low (L) when at least six studies of a single risk factor are evaluated.

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Promotes</th>
<th>Mixed</th>
<th>Inhibits</th>
<th>Evidence</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>+Nutrients (N, P)</td>
<td>84</td>
<td>10</td>
<td>6</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>+Water temperature</td>
<td>67</td>
<td>13</td>
<td>1</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>+Stratification</td>
<td>17</td>
<td>2</td>
<td>2</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>−CO2, +pH</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>M</td>
<td>H</td>
</tr>
<tr>
<td>+Precipitation</td>
<td>9</td>
<td>8</td>
<td>4</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>−Light, +turbidity</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>H</td>
<td>L</td>
</tr>
<tr>
<td>+Watershed loading</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>+Salinity</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>+Residence time</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>L &lt;6</td>
<td></td>
</tr>
<tr>
<td>+Oxygen (euphotic zone)</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>+Grazing</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>M</td>
<td>L</td>
</tr>
<tr>
<td>+COD</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>L &lt;6</td>
<td></td>
</tr>
<tr>
<td>Other (miscellaneous)</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Evidence) similar to West et al. (2012). High agreement is found for four classes of stressors: decreased CO2/increased pH, increased nutrients, increased stratification, and increased water temperature, although only the first three are rated as high evidence.

Defining risk factors for the occurrence of cyanobacterial blooms is challenging because the subject encompasses an entire phylum with hundreds of genera and thousands of known species that encompass a wide range of environmental adaptations and growth strategies. Literature on cyanobacterial blooms tends to focus on a small subset of genera that are frequent causes of excessive blooms that degrade water quality conditions and especially those genera that include potential toxin producers (e.g., species of Microcystis, Dolichospermum, Aphanizomenon). Even this subset includes vast diversity in species’ growth optima, competitive traits, etc., making general statements about cyanobacterial bloom vulnerability difficult.

The literature is in strong agreement that higher frequency of cyanobacterial blooms is associated with eutrophication due to nutrient enrichment and increased water temperature. Both of these stressor classes are likely to intensify under future climate. The responses of cyanobacteria blooms to these stressors are neither linear nor univariate, vary across genera (e.g., Rigosi et al. 2014), and can be modulated by a wide range of interacting physical, chemical, and biological factors. This has led various authors to conclude that bloom vulnerability should be evaluated on a site-specific basis, although this has not precluded widespread speculation as to how climate change might worsen problems associated with cyanobacteria. Ratios of N to P have also been identified by some as a contributor to cyanobacterial blooms, with low N systems providing an advantage to nitrogen-fixing cyanobacteria (e.g., Molot et al. 2021); however, N concentrations and N to P ratio may be more important for cyanobacterial species succession, including prevalence of toxin producers, than for ultimate cyanobacterial biomass. (We intentionally state that blooms are associated with rather than caused by eutrophication and increased temperature to emphasize the complexity and interactions among these stressors.)

Nutrients are a requirement for building algal biomass, so bloom vulnerability must be associated with nutrient availability. However, whether excess nutrients result in blooms of cyanobacteria or other plants/algae depends on a host of other factors such as temperature, mixing regime, and the timing of nutrient delivery relative to warmer temperatures that favor cyanobacteria. It appears on the one hand that further increases in nutrient loads by themselves may do little to increase the vulnerability to cyanobacterial blooms in systems that are already eutrophic. On the other hand, waterbodies with low nutrient concentrations can still produce cyanobacterial blooms in response to pulses of nutrients that are rapidly taken up and stored by cyanobacteria.

Water temperature is also a clear and known factor in vulnerability to cyanobacterial blooms but has a variety of direct and indirect effects. The most obvious linkage is that most of the major bloom-forming cyanobacteria taxa are adapted to warmer temperatures and have higher temperatures for maximum growth than common eukaryotic algae. A common rule of thumb gives the advantage to cyanobacteria when daily average water temperatures get above approximately 25°–30°C (e.g., Paerl and Huisman 2009; Yan et al. 2020; Chapra et al. 2017), although the advantage may depend on sufficient nutrients being available. The response is not linear, however, as respiration by cyanobacteria continues to increase with temperature while primary production is limited at high temperatures resulting in the inhibition of net growth above about 35°C.

There are also many other indirect effects of increased temperature on cyanobacteria, as summarized by Huisman et al. (2018) and Roussou et al. (2020), among others. One of the most significant is that warmer temperatures increase the strength and temporal extent of thermal stratification in the water column. Many cyanobacteria taxa are able to regulate their
buoyancy (through a variety of strategies) enabling them to move in the water column to retrieve nutrients from deeper levels or ascend to the surface where they can shade out competitors. A strongly stratified water column can promote hypoxia and result in increased regeneration of legacy nutrients from sediment but may also result in nutrients being sequestered in the hypolimnion, potentially reducing growth potential—unless accompanied by episodic mixing events, suggesting the non-linear complexity of the relationship. Some studies attributing cyanobacterial blooms to increased water temperature may be observing proxies for changes in the vertical mixing regime, which, although driven in part by air temperature, is also strongly affected by wind, precipitation, and the amount and timing of inflows.

Low agreement about a stressor class may be due to correlation with other factors and/or to nonlinear responses. Details are provided in the online supplemental material. For example, increased precipitation can increase bloom risk because it is associated with increased phosphorus loading; however, large runoff events can destabilize the water column, which is unfavorable to bloom formation, and flush cyanobacteria out of the system. The timing of precipitation relative to temperature and nutrient conditions favorable to cyanobacteria growth may also be important.

For decreased CO₂ and increased pH, the literature is rated as having high agreement that these factors are associated with increased vulnerability to cyanobacterial blooms. The extent to which this can be characterized as a cause versus effect is unclear: Cyanobacteria may be able to outcompete eukaryotic algae for carbon when CO₂ is in short supply, but the occurrence of cyanobacterial blooms may cause depletion of CO₂ and a concomitant increase in pH (Visser et al. 2016).

b. BRT analyses

The BRT analyses convert future climate projections (at a monthly scale) to corresponding measures of lake conditions, with a focus on surface water temperature, stratification/mixing, and the timing of these phenomena that are believed to influence the vulnerability to cyanobacterial blooms. The BRT analysis was able to provide an excellent fit to LISSS-simulated water temperatures across the 1701 lakes in the Sobol sample and a somewhat less exact fit to stratification and mixing events, which are more dependent on submonthly events. Summary results of the BRT models are shown in Table 4. These results were superior to the fit obtained from RF, with average reductions of about 38% in both RMSE and mean absolute error (MAE).

One of the key factors in increasing vulnerability to cyanobacterial blooms is likely to be an increase in summer water temperatures in the photic zone. Projected changes in average August surface water temperatures are shown in Fig. 2 and exhibit a distinct latitudinal gradient, with large changes expected across a swath of the south and north-central regions and relatively small changes for most lakes at intermediate latitudes. Differences in changes in water temperature among closely spaced lakes are primarily due to lake geometry, especially the relationship of surface area to depth.

c. Lake vulnerability analysis

Two of the key risk factors for increased cyanobacterial blooms in lakes are increased nutrient load and increased water temperatures. The interrelationship between these two and their timing is likely important as noted by several authors (Anneville et al. 2005; Elliott 2012; Rigosi et al. 2014; Scavia et al. 2021). Individually, the risks from both sources are expected to increase as shown by general increases in June–August surface water temperature (Fig. 2, above), with average temperatures of 25°C, which tend to favor cyanobacteria, occurring earlier in the year (Fig. 3), and increases in nutrient loading transported by surface runoff, which occur in most areas outside parts of the South and Southwest (Fig. 4).

We used the risk factors identified in section 4a to create a series of summary metrics that estimate projected future changes in vulnerability relative to historic conditions based on the average of MACA downscaled GCM output from the six selected GCMs. We focus on relative change from historic to future conditions in recognition of the difficulties in predicting existing bloom risks in individual lakes. The individual metrics are qualitative but allow for a visual interpretation of the distribution of relative changes in vulnerability to increased cyanobacterial blooms. For example, the literature suggests that shifts to higher summer water temperature increase the relative risk of cyanobacterial blooms even in colder lakes; however, the risk accelerates at higher temperatures (e.g., above ~25°C) where cyanobacteria gain a further competitive advantage relative to most eukaryotic algae. We therefore include a simple measure of summer water temperature increase as one risk component and added a higher

<table>
<thead>
<tr>
<th>Response</th>
<th>N</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer water temperature (°C; Jun–Aug mean)</td>
<td>1701</td>
<td>0.051</td>
<td>0.037</td>
</tr>
<tr>
<td>Days to reach 20°C (yearday)</td>
<td>1701</td>
<td>1.043</td>
<td>0.713</td>
</tr>
<tr>
<td>Days to reach 25°C (yearday)</td>
<td>1674</td>
<td>1.436</td>
<td>0.947</td>
</tr>
<tr>
<td>Days to reach 30°C (yearday)</td>
<td>824</td>
<td>1.413</td>
<td>0.844</td>
</tr>
<tr>
<td>Mixing events per year</td>
<td>1701</td>
<td>4.669</td>
<td>1.813</td>
</tr>
<tr>
<td>Stratification strength (°C)</td>
<td>1701</td>
<td>0.648</td>
<td>0.439</td>
</tr>
<tr>
<td>Date of stratification onset (yearday)</td>
<td>1673</td>
<td>5.713</td>
<td>2.225</td>
</tr>
<tr>
<td>August thermocline depth (m)</td>
<td>1701</td>
<td>0.913</td>
<td>0.441</td>
</tr>
</tbody>
</table>
ranking when this threshold moves earlier in the year. The risk metrics are described in the following paragraphs.

The *temperature risk metric* (T-RISK) is a composite risk metric associated with water temperature change in the range from 0 (low) to 3 (high). It is associated with apparent competitive advantage to cyanobacteria occurring when average surface temperatures exceed 25°C and consists of three independently additive components assigning 1 point each for 1) when average August water temperature increases by more than 3°C, 2) when first day of the year on which water temperature reaches 25°C...
occurs prior to 1 May, and 3) when day of the year on which wa-
ter temperature reaches 25°C is earlier than the historical base-
line by at least 2 weeks.

The stratification/mixing risk (S-RISK) is a composite risk
metric associated with stratification and mixing in the range
from 0 (low) to 3 (high). It is associated with proposed advan-
tage to cyanobacteria when water column is stably stratifi-
cation but thermocline depth is relatively small, and it consists of
three independently additive components assigning 1 point
each for 1) when average date of stratification onset (even if
temporary) occurs before 1 May and is more than 2 weeks
earlier than historic conditions, 2) number of simulated mix-
ing events decreases by more than 4 per year, and 3) summer
thermocline depth is less than 5 m.

The eutrophication risk metric (Eutro-RISK) is a composite
risk metric associated with nutrients and eutrophication in the
range from 0 (low) to 4 (high). It is associated with observa-
tions that risk of cyanobacterial blooms increases with in-
creased P concentration, especially when the timing is such
that the centroid of annual local surface runoff volume occurs
after average water temperature exceeds 20°C facilitating
stronger growth and competition for nutrients by cyanobacte-
reria relative to eukaryotes. It consists of the following cumula-
tive scoring levels: 0 when watershed P surplus identified in
the nutrient inventory is less than the 25th percentile of all NLA 2007 lakes, 1 when watershed P surplus identified in the
nutrient inventory is equal to or greater than the 25th percent-
ile of all NLA 2007 lakes, 2 when condition 1 is met plus the
increase in local surface runoff relative to historic conditions is
greater than 10%, 3 when condition 2 is met plus the centroid
of annual average surface runoff occurs after the average date
at which the average surface water temperature reaches 20°C,
and 4 when condition 3 is met plus the average date of the cen-
troid of annual surface runoff volume moves from before to
after the date at which the average surface water temperature
reaches 20°C.

Results of the T-RISK metric are shown for mid- and late-
century conditions in Fig. 5, while the S-RISK metric is shown
in Fig. 6 and the Eutro-RISK metric in Fig. 7. Figure 8 is a
composite risk summary measure that adds up the three pre-
ceding scores to aid in visualizing the coherence (or lack thereof) of the risk factors. The composite measure assumes
equal weighting of T-RISK, S-RISK, and Eutro-RISK. The
choice of equal weighting is a simplifying assumption. Alter-
native (unequal) weights could be assigned, but do not have a
firm basis in the summarized literature. The visualization is
thus qualitative but assumes that the different categories of
risk are synergetic.

The relative risk metric maps reveal complex temporal
and spatial patterns. Risk metrics generally increase from
midcentury to late century. The greatest T-RISK and S-RISK
metrics tend to occur across the northern tier and southern
tier of states, while Eutro-RISK is higher for the upper
Midwest and Eastern Seaboard. In interpreting the maps,
it is important to keep in mind that the analysis attempts to
identify the change in risk of cyanobacterial blooms in re-
sponse to climate change, not the probability that such
blooms will occur. Thus, there are lakes within the dataset
that already experience frequent large cyanobacterial blooms
but do not appear to be at risk of increasing vulnerability to
such blooms where the change in climate-related stressors is
small.
Cyanobacteria blooms are affected by both proximate (e.g., water quality time series, biological interactions) and ultimate (e.g., climate, morphometry, geology, land use) factors along various causal pathways. We focus here primarily on the ultimate factors and how changes in climate may interact with other factors. The risk metrics are closely tied to predicted increases in surface water temperature, and whether those temperature increases exceed certain threshold levels that have been discussed in the literature to induce an advantage for cyanobacteria versus eukaryotic algae. The risk metrics are also correlated to the physical and geographic characteristics of individual lakes shown in the NLA, which help determine

Fig. 5. T-RISK for increased cyanobacterial bloom vulnerability for (top) midcentury (circa 2040–69) and (bottom) late-century (circa 2070–99) conditions. Values shown are relative risk from low (0) to high (3).
the responses of both water temperature and stratification regime to changes in air temperature, as shown by the Pearson correlation coefficients $r$ in Table 5. The likelihood of strong stratification in a lake corresponds to a low geometry ratio.

All of the risk metrics are negatively correlated with elevation and depth and positively correlated with the geometry ratio (which is inversely dependent on depth). Lakes at higher elevation on average have greater depths and lower geometry ratios ($r$ values of 0.14 and $-0.13$, respectively), which, together with expected lower air temperatures at higher altitudes, reduce their average vulnerability to cyanobacterial blooms. The likelihood of strong stratification in a lake corresponds to a low geometry.
ratio and lakes with high geometry ratios (primarily shallow lakes) have the greatest vulnerability.

5. Discussion

Despite many decades of research, understanding of the conditions that cause cyanobacterial blooms remains limited. As summarized above, many factors are correlated with such blooms, but causal linkages remain obscure, likely reflecting the interaction of many nonlinear responses between cyanobacteria and eukaryotic algae.

Lakes and reservoirs in many parts of the United States are already experiencing problems associated with cyanobacterial
blooms and are actively involved in seeking mitigation strategies (Handler et al. 2023; Smucker et al. 2021). The focus here is on the potential change in lake and reservoir vulnerability in response to climate change to assess in what regions and for what lake characteristics is the risk due to HABs likely to increase under projected future climate. Results are intended to help identify where additional surveillance and management measures may be required to reduce or prevent future problems.

It is instructive to compare our maps of potential future risk of increased impacts of cyanobacterial blooms with the

FIG. 8. Summary risk metric for increased cyanobacterial bloom vulnerability for (top) midcentury (circa 2040–69) and (bottom) late-century (circa 2070–99) conditions. Values shown are relative risk from low (0) to high (10), although the highest predicted value is 8.
maps of existing occurrence of such blooms over the NLA dataset developed by Handler et al. (2023) based on a combination of remote sensing and NLA data. Handler et al. show that the most significant problems with existing cyanobacterial blooms are focused in the upper Midwest and Gulf Coast regions of the United States. Many lakes in these regions have limited potential for worsening bloom condition because conditions are already impaired. Our analyses are complementary to Handler et al. and suggest that currently unimpacted lakes in the interior South and Southeast and in the more western parts of the upper Midwest in particular face an increasing risk of cyanobacterial blooms.

Another point of comparison is the national study of Chapra et al. (2017), who used process-based models to predict changes in cyanobacteria concentrations in response to climate change although, in contrast to the NLA, they focus on larger, deeper waterbodies and simulate 300 individual large reservoirs, 10 natural lakes, and synthetic aggregates intended to represent smaller reservoirs and summarize the results as area-weighted averages at the four-digit hydrologic unit code (HUC) scale.

Chapra et al. note that their results are most sensitive to water temperature, nutrient concentrations, and uncertainty in cyanobacteria growth at high temperatures, all consistent with the risk factors in this paper. Their results are, however, largely predetermined by the assumptions of causal relationships that promote cyanobacteria growth built into the models.

Many of the large reservoirs addressed by Chapra et al. are also impacted by wastewater nutrient discharges and assume that such loads will increase linearly with population growth in as estimated from Integrated Climate and Land Use Scenarios (ICLUS; Bierwagen et al. 2010). (This assumption is suspect if there are water supply shortfalls under future climate, an increased focus on water reuse, and potential improvements in treatment technology.) Our study does not evaluate changes in point source discharges, although changes that lead to an increase in lake nutrient concentrations present an obvious risk of increased cyanobacterial blooms. The focus on larger reservoirs, averaging to the HUC 4 scale, and the inclusion of wastewater nutrient inputs preclude a direct comparison of the spatial patterns of cyanobacteria response to climate change in Chapra et al. (2017) with those suggested in this paper.

In considering the results presented here, we note that the occurrence of cyanobacterial blooms is a prerequisite to but not a direct indicator of the risk of high algal toxin levels. Toxin production depends on the dominance and succession of toxin producing species and the benefits of toxin production in reducing predation versus its relatively high energetic cost. Resolving those factors is likely to require detailed lake-specific analyses to fully evaluate (Tanvir et al. 2021; Howard et al. 2021; Shan et al. 2019). However, the risk metrics presented here may help in identifying where monitoring and evaluation of potential toxin problems may be warranted.

Climate projections show a range of plausible future change driven by differences in future greenhouse gas (GHG) emissions. Our analyses use the CMIP5 RCP8.5 scenario, which represents a high-end estimate of future emission rates without significant reductions. While RCP8.5 provides a pessimistic estimate of “business as usual,” it provides a good benchmark for characterizing risk and helps to provide a clear signal against a noisy background of individual lake characteristics. Evaluation of adaptation responses based on these results will need to further consider cost and risk trade-offs.

### 6. Conclusions

In this study we review the literature to identify risk factors that increase lake vulnerability to cyanobacterial blooms and evaluate how climate change may alter these factors across the sample of conterminous U.S. lakes contained in the 2007 National Lakes Assessment. Climate change is likely to increase lake vulnerability to blooms because of the associations of warmer water temperatures and a stable, stratified water column with cyanobacterial dominance. Both water temperature and stratification strength are predicted to increase throughout the twenty-first century for most lakes and for most climate scenarios. Changes in the amount and timing of watershed nutrient loads and stratification onset relative to temperature conditions that favor cyanobacteria will cause complex interactive effects in individual lakes.

The ability of process-based biogeophysical models to predict the magnitude and timing of cyanobacterial blooms is limited due to the complexity of processes and ecological succession dynamics that result in blooms. Therefore, our intent was not to predict future cyanobacterial biomass or toxin concentrations but rather to identify lakes and reservoirs that are likely to exhibit increased vulnerability to the occurrence of blooms. This was done by extracting consensus risk factors from the scientific literature and combining them with conditions derived from downscaled future climate data and a simplified model of lake thermal and mixing dynamics applied to
Results can be used to help inform climate change adaption planning to reduce the risk of HAB related impacts. Lakes that are identified as having increasing vulnerability to cyanobacterial blooms may require more intensive management efforts (if cyanobacterial blooms are already occurring) or increased surveillance in addition to management if they are not yet occurring. Short-term early warning predictors for blooms (e.g., Wilkinson et al. 2018) and remote monitoring methods (e.g., Coffer et al. 2021) currently being developed should be applied to these lakes.

This study is only a first step in understanding a complex topic. Results are conditional on the methodology developed here. Continued research is needed including detailed lake-specific analyses to fully evaluate climate change effects on HAB risk in different regional and physiographic settings.

Acknowledgments. We thank Kateri Salk-Gunderson, Anna Hamilton, and Saumya Sarkar for providing helpful discussions and review contributions. Sarkar also developed earlier versions of various Python scripts used in the analysis. Jim Carlet, Amalia Handler, and Nathan Smucker provided helpful review comments. Robert Sabo kindly shared U.S. EPA nutrient inventory results for individual NLA lakes. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 2 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Funding for this work was provided by the U.S. Environmental Protection Agency Office of Research and Development. The views expressed in this paper represent those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

Data availability statement. Data and code are archived in an online repository (see Butcher 2022).

REFERENCES


