

# Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes

Jeff C. Ho <sup>1,2\*</sup> Anna M. Michalak <sup>1</sup>

<sup>1</sup>Department of Global Ecology, Carnegie Institution for Science, Stanford, California

<sup>2</sup>Department of Civil and Environmental Engineering, Stanford University, Stanford, California

## Abstract

Climate change is expected to impact the severity of harmful algal blooms in lakes and reservoirs through a number of mechanisms related to the influence of warming temperatures and changes to precipitation patterns. Evidence on the prevalence of individual mechanisms is lacking, however, with knowledge of many mechanisms restricted to studies of individual or small subsets of lakes. Here, we leverage over twelve hundred summertime lake observations from across the continental U.S. to explore evidence for the hypothesized risks from climate change attributable to specific mechanisms. Using a statistical model selection approach, we examine associations between temperature and precipitation variables and indicators of total phytoplankton abundance, species dominance, and toxicity. We find evidence in support of the hypotheses that summer temperatures drive total abundance, that the length of the summer drives cyanobacterial abundance, and that increased temperatures may reduce the observed toxicity of blooms in some cases. We find that nutrient concentrations are also likely to be impacted by lake warming, as increased temperatures are robustly associated with increased total phosphorus concentrations. Evidence for the impact of precipitation is mixed, however, as there is evidence to support that increased nutrient runoff from precipitation could support blooms but also that nutrient concentrations could be reduced through greater flushing due to precipitation. While statistical associations are not definitive evidence of formal mechanistic links, the geographic scale of the results is useful for identifying hypothesized mechanisms that are widespread across the continental U.S., and therefore for informing understanding of the influence of climate change.

Harmful algal blooms (HABs) are hypothesized to be increasing globally (Paerl and Paul 2012), exacerbated by changing climatic conditions (Paerl and Huisman 2009). Freshwater HABs impact aquatic food production, recreation and tourism, and drinking water supplies, leading to economic losses of \$4.6 billion annually in the United States alone (Kudela et al. 2015). HABs dominated by toxic cyanobacterial species are especially concerning due to their potential impact on human health, having been linked to incidents of liver damage and severe skin irritation (Chorus and Bartram 1999; Geh et al. 2016). While there are a number of factors influencing the frequency of HABs, climate change has emerged as one of the most important as it is potentially counteracting local management efforts to improve bloom conditions (Winder 2012). Understanding the potential effects of climate change on HABs in freshwater

ecosystems is therefore imperative for safeguarding current and future water quality (Michalak 2016).

Climate change is hypothesized to impact the proliferation of HABs in lakes and reservoirs through a number of mechanisms (Paerl and Huisman 2009; Visser et al. 2016; Huisman et al. 2018). First, increased temperature is hypothesized to lead to selection for more harmful species. This occurs through increased growth rates of harmful cyanobacterial species relative to other species (Paerl and Huisman 2008; Carey et al. 2012; Visser et al. 2016), through extending of the summer growing season (Anneville et al. 2005; Deng et al. 2014; Visser et al. 2016) and through increased vertical stratification resulting in greater fitness by cyanobacteria that are able to regulate their buoyancy (Wagner and Adrian 2009; Carey et al. 2012; Posch et al. 2012; Taranu et al. 2012; Visser et al. 2016). Higher temperature also increases internal phosphorus loading from sediments through increased rates of soil diffusivity and increased rates of soil phosphorus regeneration (Matisoff et al. 2016), and this internal phosphorus source has been shown be a significant proportion of the overall phosphorus budget in some lakes (Jeppesen et al. 2007; Ho and Michalak 2017; Orihel et al. 2017; Ding et al. 2018). Second, greater overall

\*Correspondence: jeffho@alumni.stanford.edu

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Additional Supporting Information may be found in the online version of this article.

precipitation and more extreme precipitation events are hypothesized to increase HABs through the generation of greater nutrient runoff (Reichwaldt and Ghadouani 2012; Michalak et al. 2013), although this could be balanced by greater precipitation flushing nutrient-rich waters out of the ecosystem prior to bloom formation (Anderson et al. 2002; Reichwaldt and Ghadouani 2012). Despite this broad literature, evidence on the prevalence of individual mechanisms is lacking, with knowledge of many mechanisms restricted to studies of individual, or small subsets of, lakes (Posch et al. 2012; Descy et al. 2016).

An improved understanding of the prevalence of these hypothesized mechanisms across a large number of waterbodies would make it possible to focus mitigation efforts on those mechanisms that are most widespread. Indeed, different mitigation approaches target different mechanisms: for example, artificial mixing approaches aim to remove the fitness advantage of buoyant cyanobacteria by disrupting vertical stratification (Paerl et al. 2016). Some mitigation strategies may also be more effective at targeting specific aspects of blooms. For instance, as the abundance, species dominance and toxicity of HABs may each respond differently to changes in climate (e.g., higher temperature may promote faster growing but less toxic cells; Peng et al. 2018). As both statistical and mechanistic models are already being used to project the influence of climate change on future HAB occurrence (Beaulieu et al. 2013; Chapra et al. 2017; Lin 2017), deeper knowledge of the mechanisms underlying climatic impacts on HABs is critical for informing such models and improving the robustness of future projections. This is best accomplished by assessing which meteorological conditions, and in what combinations, make HABs most likely across a large number of lakes (Michalak 2016).

Past studies examining large numbers of lakes have primarily explored simple relationships with temperature (Kosten et al. 2012; Beaulieu et al. 2013; Rigosi et al. 2014, 2015; Mantzouki et al. 2018), and close to none have examined relationships with other predictors such as precipitation and stratification, or explored other types of relationships with temperature such as the impact of longer summers. The focus on temperature has been warranted, as increasing evidence suggests lakes are warming at a faster rate than the surrounding air (O'Reilly et al. 2015). However, the paucity of studies examining precipitation impacts is an important omission given its role in impacting nutrient runoff and residence times (Reichwaldt and Ghadouani 2012). To our knowledge, only one past study has quantified the role of precipitation on HABs across a large number of lakes (Lin 2017), and this study only examined total annual amount and intensity of precipitation, not extreme precipitation nor timing of precipitation. The latter two have been documented elsewhere as important for driving HABs and nutrient runoff (Michalak et al. 2013; Sinha and Michalak 2016). Past studies have also not explored the role of precipitation *in conjunction* with nutrient inputs, which is important because the synergistic effects of high precipitation and high

terrestrial phosphorus supply are substantial (Motew et al. 2018). Moreover, past studies have also not explored the impact of meteorology on in-lake nutrient concentrations themselves, only on external nutrient loading into waterbodies (e.g., Sinha and Michalak 2016). This distinction is important because nutrient concentrations have been shown to be the strongest predictors of HABs (Beaulieu et al. 2013; Rigosi et al. 2014; Taranu et al. 2017), and many factors influence the relationship between external nutrient loading and nutrient concentrations (e.g., residence time). Testing specific hypotheses about how climate change is expected to impact internal phosphorus loading furthermore requires examining in situ nutrient concentrations rather than external nutrient loads.

Here, we leverage two country-wide lake surveys along with meteorological data across the continental U.S. to explore whether relationships between meteorology and HABs can be detected in lakes. We explore relationships between HABs and temperature, precipitation, and other variables representing their combined impacts. The statistical importance of different relationships is used to inform possible prevalence of underlying mechanisms across a geographically diverse set of lakes, and the implications for hypothesized climatic impacts on HABs are discussed. In this work, we aim to answer two main questions:

1. What best explains summertime phytoplankton bloom abundance, cyanobacterial dominance, and toxicity in lakes across the continental U.S. when examining the effects of longer summers, thermal stratification, and temperature directly, as well as the effects of precipitation-driven nutrient loading and flushing?
2. Based on the importance of potential meteorological drivers, which of the hypothesized mechanisms of climatic impact are most supported and/or most likely to be widespread?

## Methods

### Data sets used

We analyzed 2066 observations taken from over 1500 lakes and reservoirs across the continental United States (CONUS) selected by the U.S. Environmental Protection Agency (EPA) for the 2007 and 2012 National Lake Assessment (NLA) surveys. Although the latter survey was performed in 2012, the resultant data became available only in 2017 and thus far few studies have examined data from both surveys together (e.g., Leech et al. 2018). A total of 1028 and 1038 observations were taken in 2007 and 2012, respectively (U.S. Environmental Protection Agency 2010, 2016). Each lake or reservoir was visited once or twice between May and October for each year of the survey (2007 and/or 2012). Survey lakes were selected based on a state by lake size stratified probabilistic sampling design in order to be representative of the roughly 50,000 lakes in the CONUS (Peck et al. 2013). Because only 401 lakes were visited in both 2007 and 2012 (Leech et al. 2018), we combined observations from the two surveys to maximize the number of lakes with available data, which was made possible

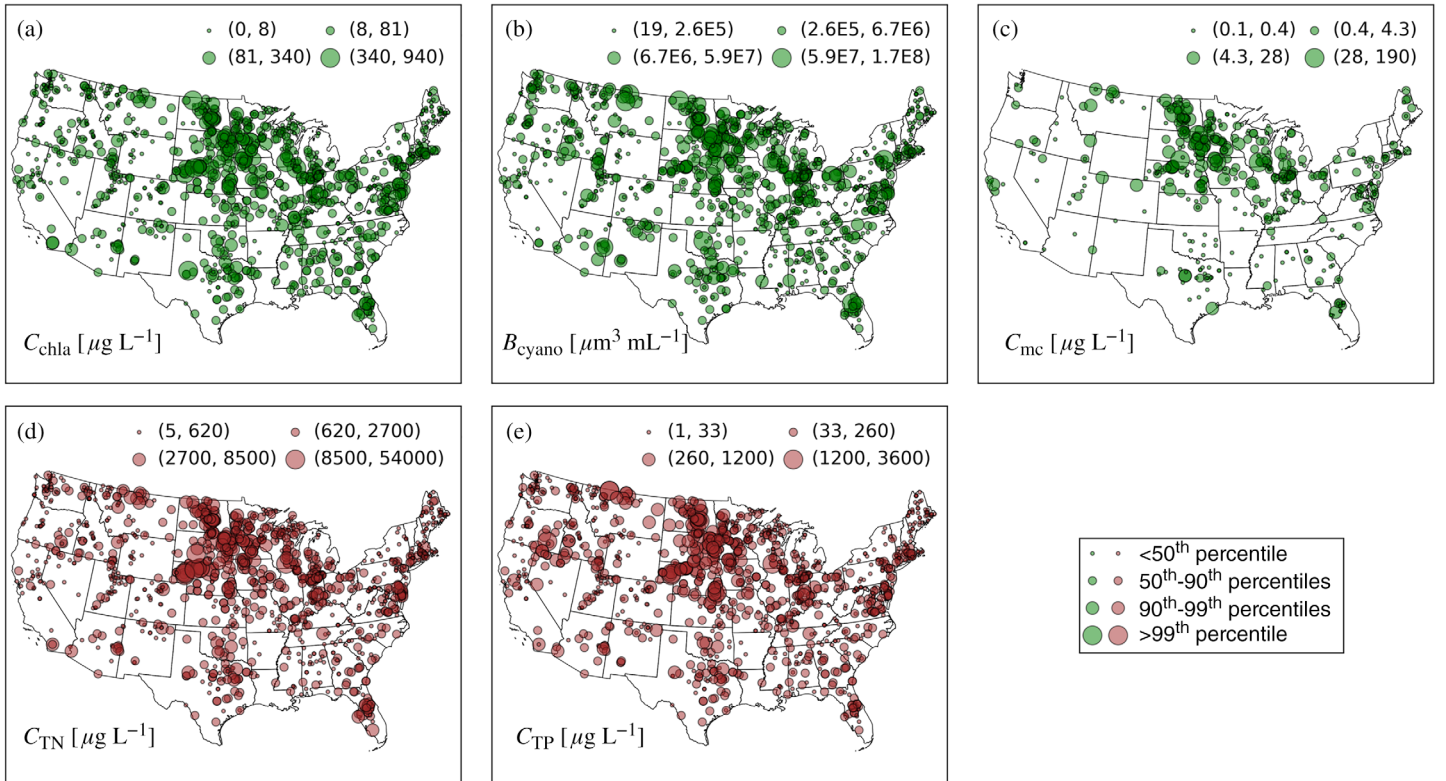
by the fact that the field and analysis protocols for the variables considered here was equivalent across the two surveys. All data are freely available online (<http://www.epa.gov/national-aquatic-resource-surveys>).

From these data sets, we selected as response variables three variables related to HABs, namely chlorophyll *a* ( $C_{chl a}$ ), cyanobacterial biovolume ( $B_{cyano}$ ), and microcystin concentration ( $C_{mc}$ ), and two related to nutrient status, namely in situ total nitrogen concentration ( $C_{TN}$ ) and total phosphorus concentration ( $C_{TP}$ ). The geographic distributions of these five variables across the CONUS are presented in Fig. 1. Using these data enabled comparison to many past studies that have examined the 2007 data set (Beaulieu et al. 2013; Rigosi et al. 2014; Lin 2017; Taranu et al. 2017). The data collection procedures and sampling regime for these data have been described at length in prior studies and in official documentation (U.S. Environmental Protection Agency 2007, 2011).

For microcystin concentration, only values above the detection limit were used (see Table 1). A full, zero-adjusted statistical approach taking into account both the factors affecting detection as well as severity once observed (Taranu et al. 2017) is out of the scope of this study's exploration of the relationships between meteorology and HABs. To our knowledge,

there have not been any hypothesized mechanisms regarding temperature or precipitation influencing the detectability of microcystins. There have, however, been several studies providing hypothesized mechanisms on the impact of temperature on toxin concentrations (e.g., from both laboratory studies, Peng et al. 2018, and taxonomic studies, Anneville et al. 2005). Using only microcystin values above the detection limit therefore best addresses the research questions in this study, and provides comparability to the second part of the hurdle model described by Taranu et al. (2017).

In contrast to past studies, we focused on a subset of observations collected during the two warmest months of the year (July and August,  $n = 1260$ ) rather than the full set of observations collected from May to October. This is because values of  $C_{chl a}$ ,  $B_{cyano}$ , and  $C_{mc}$  were highest in July and August, and therefore better represent peak HAB severity and the cumulative impact of nutrient loading and meteorological conditions. Examining summary statistics for the July–August subset (Table 1) suggests a comparable range in values for HAB indicator, nutrient concentration, and lake characteristic variables as reported in previous studies (Beaulieu et al. 2013; Taranu et al. 2017), with slightly higher mean values of HAB indicator variables, as expected.



**Fig. 1.** Geographic distribution of HAB indicators in green, including (a) Chl *a* concentration ( $C_{chl a}$ ), (b) cyanobacterial biovolume ( $B_{cyano}$ ) and (c) microcystin concentration when above the detection limit ( $C_{mc}$ ), and in situ (d) total nitrogen ( $C_{TN}$ ) and (e) total phosphorus ( $C_{TP}$ ) concentrations in maroon. Each circle represents one measurement from either the NLA 2007 or NLA 2012 data sets, up to 1260 in total (see “Methods” section). The size of the circle represents the magnitude of the observation, indicated by the legend.

**Table 1.** Summary statistics of key variables across the 1260 lake observations used in this study. Observations are obtained from both the NLA 2007 and 2012 lake surveys (U.S. Environmental Protection Agency 2010, 2016), selecting only those observations taken in July and August.

Variable		Min	Max	Mean	Median	<i>n</i> , Total	<i>n</i> , above detection limit
HAB indicators	Chl <i>a</i> ( $\mu\text{g L}^{-1}$ )	0	940	32	8	1254	1254
	Cyanobacterial biovolume ( $\mu\text{m}^3 \text{mL}^{-1}$ )	19	$1.7 \times 10^8$	$3.1 \times 10^6$	$2.6 \times 10^5$	1208	1208
	Microcystin ( $\mu\text{g L}^{-1}$ )	0.1	190	2.3	0.4	1260	496
Nutrient concentrations	TN ( $\mu\text{g L}^{-1}$ )	5	54,000	1270	620	1260	1260
	TP ( $\mu\text{g L}^{-1}$ )	1	3600	110	33	1260	1260
Lake characteristics	Area (ha)	1	58,000	740	50	1260	N/A
	Depth (m)	0.5	60.3	8.4	5.2	1260	N/A
	Surface-water temperature ( $^{\circ}\text{C}$ )	10	35	25	25	1260	N/A
	Mean column water temperature ( $^{\circ}\text{C}$ )	7.0	34	22	23	1260	N/A

Ancillary data obtained from the National Lakes Assessment data sets included water temperature profiles, surface area, maximum depth, latitude, longitude, and the date of sample collection. Water temperature was measured at either 0.5 m, 1.0 m, or 2.0 m vertical intervals at the deepest point of the lake for lakes with depths less than 3 m, 3–20 m, or greater than 20 m, respectively (U.S. Environmental Protection Agency 2007, 2011). Lake maximum depth and the above-mentioned response variables were also measured at the deepest point of the lake (U.S. Environmental Protection Agency 2007, 2011). Lake latitude, longitude, and surface area within the NLA data sets were obtained from the National Hydrology service data set (U.S. Geological Survey, 2007–2014, National Hydrography Dataset available on the World Wide Web, <https://nhd.usgs.gov>).

We focused on meteorological data with contiguous spatial and temporal coverage in 2007 and 2012 to obtain observations colocated and coincident with NLA observations. Daily precipitation data were obtained from Daymet Version 3 (Thornton et al. 2017) and Palmer drought severity index (PDSI) data at 10-d intervals were obtained from the University of Idaho (Abatzoglou 2013; Abatzoglou et al. 2014) over each lake watershed. Hourly air temperature data over each lake's surface were obtained from Phase 2 of the North American Land Data Assimilation project (NLDAS2; Xia et al. 2009, 2012). Computations were performed in Google Earth Engine (Gorelick et al. 2017) with lake surface and watershed boundaries obtained from the Environmental Protection Agency (U.S. Environmental Protection Agency 2010, 2016).

We also included nutrient input and loading variables based on net anthropogenic nitrogen and phosphorus inputs (NANI and NAPI, respectively). NANI is based on fertilizer nitrogen input, atmospheric deposition, agricultural nitrogen fixation, net food and feed import, and nonfood crop export (Swaney et al. 2018a,b). NAPI is similar, but without the atmospheric deposition component (Swaney et al. 2018a). Both were

calculated for 2007 and 2012 using the NANI and NAPI toolboxes Version 3.1 (<http://www.eeb.cornell.edu/biogeonanc/nani/nani.htm>). In addition, using 2006 and 2011 land cover data from the U.S. National Land Cover database (Fry et al. 2011; Homer et al. 2015), in combination with the previously mentioned precipitation data, we calculated total nitrogen (TN) loading estimates at the Hydrologic Unit Code 8 (HUC8) watershed scale based on a previously published statistical model (Sinha and Michalak 2016). These TN loading values were then aggregated to study-lake watersheds by multiplying the proportion of each HUC8 watershed within each lake watershed by its TN loading and summing for each lake watershed.

### Model development

We used variables related to precipitation, air and water temperature, stratification, drought severity, and geomorphology, in addition to nutrient input and loading variables, to develop linear models for response variables, calibrated using up to 1260 summer lake observations. HAB indicators and nutrient concentration variables were natural-log transformed to adjust for skewness and to provide a clearer comparison to past studies (Beaulieu et al. 2013; Rigosi et al. 2014; Taranu et al. 2017). Response variables included  $\ln(C_{\text{chl}a})$ ,  $\ln(B_{\text{cyano}})$ , and  $\ln(C_{\text{mc}})$ , representing phytoplankton abundance, cyanobacterial dominance, and toxicity, respectively, and  $\ln(C_{\text{TN}})$  and  $\ln(C_{\text{TP}})$ , representing nutrient status. The models took the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1)$$

where  $\mathbf{y}$  ( $n \times 1$ ) represents observations of each response variable, for example,  $\ln(C_{\text{mc}})$ ,  $\mathbf{X}$  ( $n \times [k + 1]$ ) is a matrix of predictor variables and a column of ones representing the intercept term,  $\boldsymbol{\beta}$  ( $[k + 1] \times 1$ ) is a vector of drift coefficients and the intercept,  $\boldsymbol{\epsilon}$  ( $n \times 1$ ) is a vector of residuals,  $k$  represents the number of predictor variables in a particular model (excluding

**Table 2.** Best models determined by BIC for each response variable, showing coefficient of determination ( $R^2$ ) and number of parameters ( $k$ ) for each best model. Adjusted  $R^2$  are identical to the  $R^2$  values to two significant digits. All regression coefficients are statistically significant at the 0.001 level based on standard  $t$ -tests.  $C_{\text{chla}}$  represents Chl  $a$  concentration ( $\mu\text{g L}^{-1}$ );  $B_{\text{cyano}}$  represents cyanobacterial biovolume ( $\mu\text{m}^3 \text{mL}^{-1}$ );  $C_{\text{mc}}$  represents microcystin concentration when detected ( $\mu\text{g L}^{-1}$ );  $C_{\text{TN}}$  represents total nitrogen concentration ( $\mu\text{g L}^{-1}$ ); and  $C_{\text{TP}}$  represents total phosphorus concentration ( $\mu\text{g L}^{-1}$ ).  $T_{\text{air,MAM}}$  and  $T_{\text{air,JJA}}$  represent air temperature averaged in the months of March, April, and May (MAM), and June, July, and August (JJA), respectively ( $^{\circ}\text{C}$ ).  $P_{\text{annual}}$  represents total annual precipitation (mm).  $P_{\text{JJA},p>0.95}$  represents total precipitation (mm) in JJA months on days with precipitation greater than the 95<sup>th</sup> percentile, calculated based on 30 yr of daily precipitation from 1981 to 2010.  $D$  represents depth (m) and Lon represents longitude ( $^{\circ}$ ). FNI and FPI represent nitrogen (N) and phosphorus (P) inputs from fertilizer ( $\text{kg-N km}^{-2} \text{yr}^{-1}$ ;  $\text{kg-P km}^{-2} \text{yr}^{-1}$ ).  $\ln()$  indicates natural log transform. See “Methods” section for source data and Supporting Information Table S1 for a full list of all considered variables.

	Response variable	Best model determined by BIC	$R^2$	$k$
HAB indicators	$\ln(C_{\text{chla}})$	$0.60 \ln(C_{\text{TN}}) + 0.43 \ln(C_{\text{TP}}) + 0.066 T_{\text{air,JJA}} + 0.53 \ln(P_{\text{annual}}) - 8.3$	0.69	4
	$\ln(B_{\text{cyano}})$	$0.72 \ln(C_{\text{TN}}) + 0.54 \ln(C_{\text{TP}}) + 0.059 T_{\text{air,MAM}} + 0.40 \ln(D) + 4.4$	0.28	4
	$\ln(C_{\text{mc}})$	$0.67 \ln(C_{\text{TN}}) - 0.073 T_{\text{air,MAM}} - 4.5$	0.25	2
Nutrient concentrations	$\ln(C_{\text{TN}})$	$0.00046 \text{ FPI} + 0.095 \ln(\text{FNI}) - 0.00093 P_{\text{annual}} + 0.0022 P_{\text{JJA},p>0.95} - 0.55 \ln(D) + 7.3$	0.58	5
	$\ln(C_{\text{TP}})$	$0.0010 \text{ FPI} + 0.076 T_{\text{air,JJA}} - 0.58 \ln(P_{\text{annual}}) - 0.79 \ln(D) - 0.016 \text{ Lon} + 5.2$	0.57	5

the intercept term), and  $n$  varies depending on the response variable (see Table 2).

We used a statistical model selection approach based on the Bayesian information criterion (BIC; Schwarz 1978) to select variables to be included in the linear models based on their ability to explain response variable variability. BIC measures model fit using the residual sum of squares, with model complexity penalized based on the number of predictor variables in the model:

$$\text{BIC} = -2\ln\left(\frac{(\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}})}{n}\right) + (k+1)\ln(n) \quad (2)$$

where  $\hat{\mathbf{y}}$  ( $n \times 1$ ) additionally represents the predicted values of the response variable based on a specific subset of predictor variables and  $T$  indicates a matrix transpose operation.

We chose BIC over other criterion-based model selection approaches (e.g., the Akaike information criterion) because BIC has a larger penalty term and therefore tends to select smaller (i.e., more parsimonious) sets of predictor variables (e.g., Raftery 1995). Such an approach is better suited for these data than “greedy” model selection approaches (e.g., forward/backward substitution, or single decision trees) because of the high degree of collinearity among candidate variables (Graham 2003). Unlike with such “greedy” model selection approaches, the presence of collinearity among candidate variables does not affect which variables get selected in the best model using BIC, because (1) a model with colinear predictors will only be selected if the additional variance explained overcomes the penalty term and (2) because all possible admissible combinations of candidate variables (see details below) are considered rather than only considering nested models. An initial survey of alternate approaches to handle collinearity (e.g., principal components analysis) increased model complexity without improving model

explanatory power, as did other statistical approaches from the literature (e.g., boosted regression trees; Lin 2017; Taranu et al. 2017). All possible combinations of candidate variables were considered, except when noted otherwise below, with a maximum model size of  $k = 5$  based on maximizing the robustness of the results and minimizing the risk of overfitting. The model with the lowest BIC score is typically selected as the “best” model; however, we used a BIC difference of 10 as the threshold for choosing a larger model with a lower BIC over a smaller model with a higher BIC. This threshold corresponds to “very strong” evidence in favor of the larger model (Raftery 1995), and was selected because the large sample sizes considered here ( $n$  up to 1260) mean that larger models could be favored via BIC even for very small increases in  $R^2$ .

### Candidate variables

For the HAB indicator response variables, in situ TN and total phosphorus (TP) concentrations were included as candidate variables, as well as natural-log transformations of these variables. Here, and in cases below, natural-log transformations were considered for candidate variables where values varied by several orders of magnitude. We restricted model selection to include at most one in situ TN variable and at most one in situ TP variable.

Six candidate variables were considered based on in situ water temperature and whole-lake average air temperature. Surface, bottom, and mean column water temperature observed at time of sampling were included. Annual, spring, and summer average air temperature were included as well, with spring defined as during the months of March, April, and May (MAM) and summer defined as during the months of June, July, and August (JJA). In situ water temperature has been included in many studies (Kosten et al. 2012; Beaulieu et al. 2013; Rigosi et al. 2014; Lin 2017), and air temperature has also been found to predict trends in cyanobacterial



pigment concentration (Taranu et al. 2015). We restricted model selection to include at most one temperature variable.

Fifteen candidate variables were considered based on climate change indices developed by the Expert Team on Climate Change Detection and Indices (ETCCDI, [http://etccdi.pacificclimate.org/list\\_27\\_indices.shtml](http://etccdi.pacificclimate.org/list_27_indices.shtml)). These included total annual precipitation, total precipitation during spring (MAM) and summer (JJA) months, as well as natural-log transforms of these three variables, plus nine additional candidate variables indicative of extreme precipitation. The extreme precipitation variables included extreme precipitation amounts annually and during MAM and JJA, with “extreme” defined based on specific percentiles of daily precipitation, allowing multiple percentiles as candidate variables (see Supporting Information Table S1). This approach follows others who have linked precipitation drivers with nutrient loads (Sinha and Michalak 2016). We restricted model selection to include at most one total precipitation variable and at most one extreme precipitation variable.

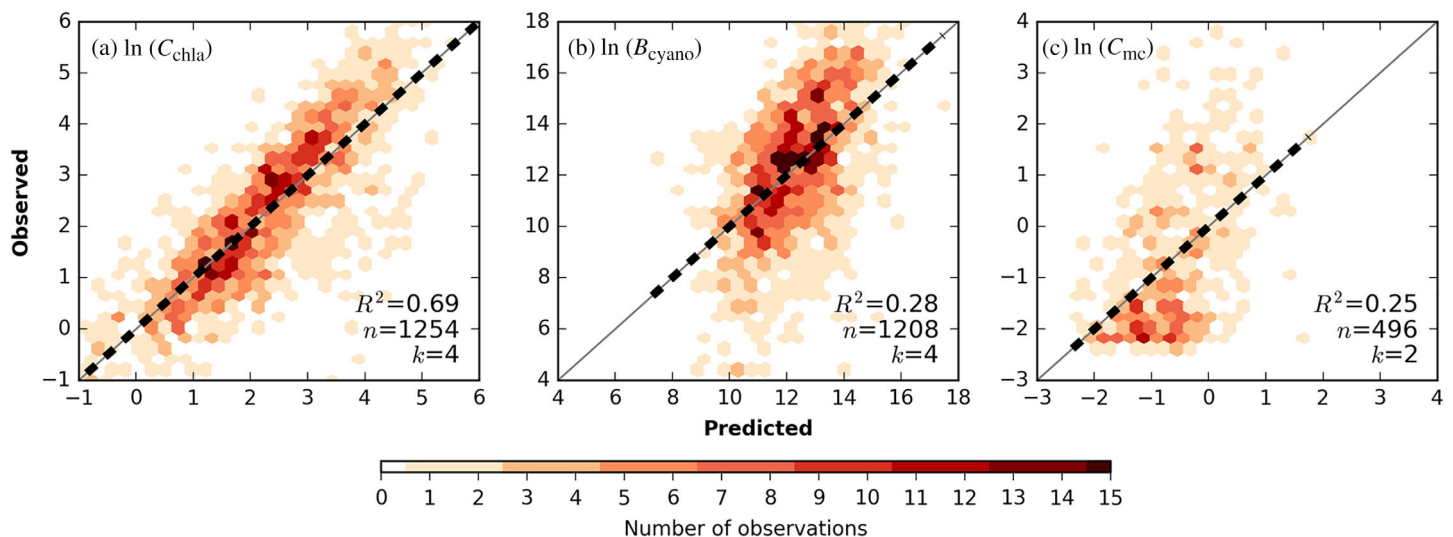
We also included variables associated with lake geomorphological characteristics and time of sampling observation to assess the degree to which climatic variables were serving as proxies for these. These included the area and the depth of the lake plus natural-log transforms of these two variables, lake latitude and longitude, and the day of year and month of lake observation as candidate variables. At most two variables from this category were allowed in model selection.

Two additional categories of nutrient-related variables (nutrient inputs and nutrient loading) were included as candidate variables. Annual NANI, NAPI, and nitrogen and phosphorus fertilizer inputs were included, with the fertilizer components considered because of previous importance in predicting cyanobacterial blooms across the CONUS (Marion et al. 2017). TN loading estimates as modeled based on HUC8

watersheds across the CONUS (Sinha and Michalak 2016) were also included (see Supporting Information Table S1). Natural-log transformed versions of nitrogen and phosphorous from fertilizer, and modeled TN loading were also considered, as were inverse hyperbolic sine transformed versions of NANI and NAPI, similarly to Sinha and Michalak (2016). At most two variables were allowed in model selection from these categories.

Three more candidate variables combining the effects of temperature and precipitation, PDSI averaged annually, spring (MAM), and summer (JJA), were also included. The rationale for including these variables is to assess the impact of the record-setting drought in 2012 (Zhou et al. 2015) on HAB indicators and nutrient concentrations, as drought intensity has been found to influence nitrogen and phosphorus limitation and subsequently the competition between different types of phytoplankton species (Hayes et al. 2015). Negative values in the PDSI data used here indicate more drought while positive values indicate less drought (Abatzoglou 2013).

Finally, four candidate variables representing the degree of stratification were also included. These included the depth of the thermocline and its natural-log transform, buoyancy frequency at thermocline depth, and the difference between the surface and bottom water temperatures. The first two are well-known indicators of stratification and were calculated using the equations described in *LakeAnalyzer* (Read et al. 2011). Thermocline depth is the depth of the maximum density gradient in the water column, controlling the depth niches of aquatic organisms (Cantin et al. 2011), and has been shown to influence the selection of harmful phytoplankton species (Posch et al. 2012). Buoyancy frequency is the angular frequency at which a parcel of water would oscillate if displaced from the water column, a measure of the strength of stratification, and influences the degree of resistance to vertical mixing



**Fig. 2.** Overall fit of best models for each HAB response variable: (a)  $\ln(C_{\text{chla}})$ , (b)  $\ln(B_{\text{cyano}})$ , and (c)  $\ln(C_{\text{mc}})$ . Predicted values are based on best models selected through BIC (see Table 2). Dashed black line indicates best fit line while gray line indicates 1:1 correspondence between observed and predicted values.

(Kraemer et al. 2015). At most one stratification or drought severity variable was included in model selection.

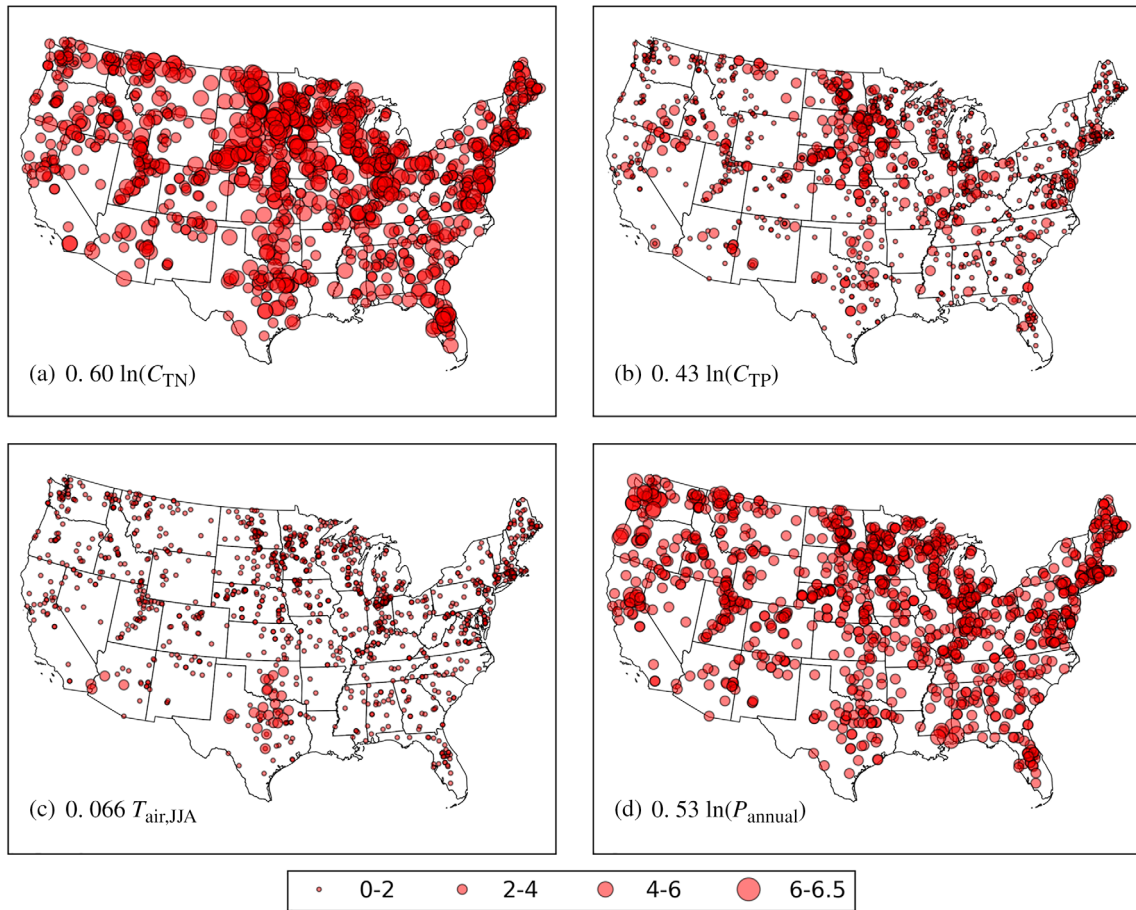
### Model interpretation and sensitivity analyses

We used Bayesian model averaging (BMA; Raftery et al. 1997) to assess the relative importance of different candidate variables in explaining variability in HAB indicators and nutrient concentrations. BMA estimates posterior probabilities of individual variables based on BIC scores of models including that variable (Yadav et al. 2010). The posterior probability is a measure of the importance of a candidate variable in explaining the observed variability in the response variable, when accounting for the considered subsets of candidate variables (Yadav et al. 2010). A variable with higher probability from BMA indicates that models that include the variable tended to have higher explanatory power and stronger evidence in favor of that model (Raftery 1995), and therefore

variables with higher probability correspond to greater importance as explanatory variables.

Though typically BMA is used to differentiate between individual variables (Ho and Michalak 2017; Del Giudice et al. 2018), we used BMA here to assess the relative importance of different *categories* of candidate variables. The posterior probability of categories of candidate variables was calculated similarly as for individual variables, except that it considered all models which include any candidate variable in a specific category rather than all models which include a specific variable (Yadav et al. 2010). In the same way as for individual variables, categories with higher probability from BMA correspond to greater importance, as higher probability indicates that models including variables from that category tended to have higher explanatory power. See Supporting Information Table S1 for which candidate variables were considered in each category.

We further used the probabilities from BMA to assess the robustness of the signs of drift coefficients of individual



**Fig. 3.** Geographic distribution of the contributions of each predictor variable to predicted log of Chl *a* concentration,  $\ln(C_{\text{chl}a})$ , based on the model in Table 2. Each circle represents one location from either the NLA 2007 or NLA 2012 data sets, 1254 in total (see Table 1). The size of each circle represents the magnitude of the contribution of each term in the model, including the contribution of (a) natural log-transformed TN concentration,  $\ln(C_{\text{TN}})$ , (b) natural log-transformed TP concentration,  $\ln(C_{\text{TP}})$ , (c) air temperature averaged over June, July, and August,  $T_{\text{air,JJA}}$  and (d) natural log-transformed total annual precipitation,  $\ln(P_{\text{annual}})$ . All predictor variables are minimum-deviated to better represent the geographic and across-term variability in contributions.

variables. For example, for a given candidate variable such as total annual precipitation, we computed the posterior probability when considering all models where total annual precipitation was included and had a positive drift coefficient as well when considering all models where total annual precipitation was included and had a negative drift coefficient. Because higher probability from BMA indicates that the considered models tend to have higher explanatory power, these probabilities can be used to assess whether models with positive or negative drift coefficients are more strongly supported by the available observations.

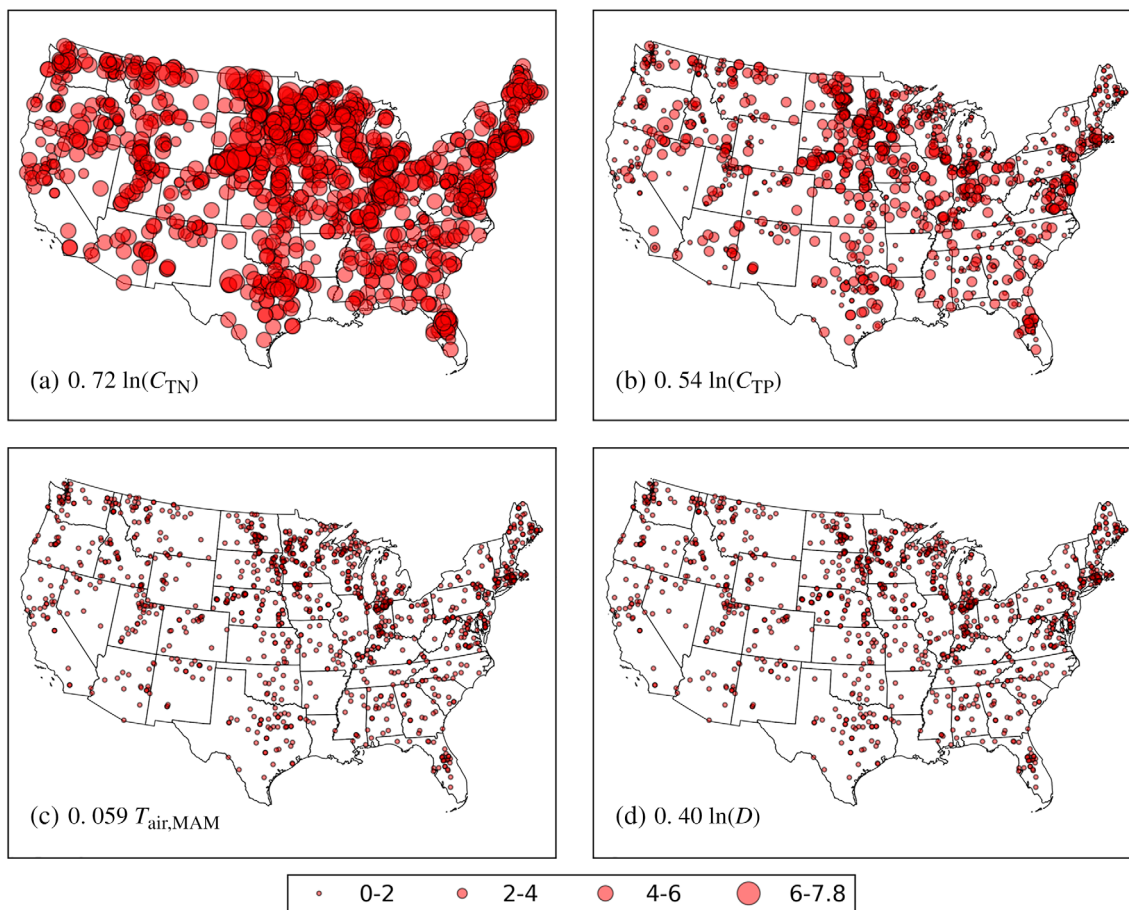
### Hypotheses for impact of climate change

When using the results of model selection to inform support for hypotheses of climatic influence on HABs, we consider the following hypotheses:

1. *Direct growth*: Higher temperatures increase cyanobacterial growth rates (Paerl and Huisman 2008; Carey et al. 2012; Visser et al. 2016). Support for this hypothesis occurs if

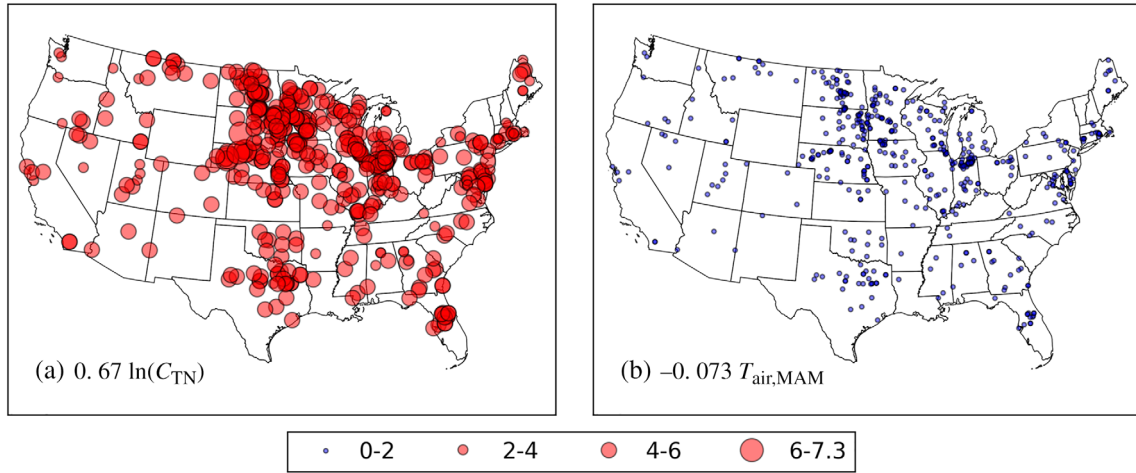
water or summer air temperature is an important predictor of HABs with a positive drift coefficient.

2. *Longer summer*: Longer seasons of high temperature increase the opportunity for cyanobacterial growth (Anneville et al. 2005; Deng et al. 2014; Visser et al. 2016). Support for this hypothesis occurs if spring air temperature is an important predictor of HABs with a positive drift coefficient.
3. *Internal phosphorus loading*: Higher temperature increases the rate of internal phosphorus loading, driving HABs (Jeppesen et al. 2007; Matisoff et al. 2016; Orihel et al. 2017; Ding et al. 2018). Support for this hypothesis occurs if water or summer air temperature is an important predictor of TP with a positive drift coefficient, *and* TP is an important predictor of HABs with a positive drift coefficient.
4. *Load driving*: Greater precipitation amounts increase nutrient loading, driving HABs (Reichwaldt and Ghadouani 2012; Michalak et al. 2013). Support for this hypothesis occurs if precipitation is an important predictor of nutrient concentrations with positive drift coefficients, *and* nutrient concentrations



**Fig. 4.** Geographic distribution of the contributions of each predictor variable to predicted log of cyanobacterial biovolume,  $\ln(B_{\text{cyano}})$ , based on the model in Table 2. Each circle represents one location from either the NLA 2007 or NLA 2012 data sets, 1208 in total (see Table 1). The size of each circle represents the magnitude of the contribution of each term in the model, including the contribution of (a) natural log-transformed TN concentration,  $\ln(C_{\text{TN}})$ , (b) natural log-transformed TP concentration,  $\ln(C_{\text{TP}})$ , (c) air temperature averaged over March, April, and May,  $T_{\text{air,MAM}}$ , and (d) natural log-transformed depth,  $\ln(D)$ . All predictor variables are minimum-deviated to better represent the geographic and across-term variability in contributions.





**Fig. 5.** Geographic distribution of the contributions of each predictor variable to predicted log of microcystin concentration (at locations where the measured concentration was above the detection limit),  $\ln(C_{mc})$ , based on the model in Table 2. Each circle represents one location from either the NLA 2007 or NLA 2012 data sets, 496 in total (see Table 1). The size of each circle represents the magnitude of the contribution of each term in the model, including the contribution of **(a)** natural log-transformed TN concentration,  $\ln(C_{TN})$  and **(b)** air temperature averaged over March, April, and May,  $T_{air,MAM}$ .  $\ln(C_{TN})$  was minimum-deviated while  $T_{air,MAM}$  was maximum-deviated (due to its negative drift coefficient; Eq. 1, Table 2) to better represent the geographic and across-term variability in contributions. Red circles indicate an increase in predicted  $\ln(C_{mc})$  while blue circles indicate a decrease in predicted  $\ln(C_{mc})$ .

are important predictors of HABs with positive drift coefficients.

5. *Flushing*: Greater precipitation amounts decrease HABs by reducing residence times of nutrients (Anderson et al. 2002; Reichwaldt and Ghadouani 2012). Support for this hypothesis occurs if precipitation is an important predictor of nutrient concentrations with *negative* drift coefficients, and nutrient concentrations are important predictors of HABs with positive drift coefficients.
6. *Stratification*: Greater vertical stratification selects for cyanobacteria (Wagner and Adrian 2009; Carey et al. 2012; Posch et al. 2012; Taranu et al. 2012; Visser et al. 2016). Support for this hypothesis occurs if any stratification variable is an important predictor of HABs with a positive drift coefficient, or if any drought severity variable (i.e., through low precipitation and high temperature) is an important predictor of HABs with a negative drift coefficient (as higher PDSI indicates wetter conditions).

We note that some of the above are competing hypotheses, and we therefore do not expect all hypotheses to be equally supported by observations. Rather, each of the above hypotheses is supported by studies of individual, or small sets of, lakes, and the goal here is to explore which of these hypotheses are more broadly supported across CONUS lakes.

## Results

The model selection approach identified environmental covariate combinations that explained 69%, 28%, and 25% of  $\ln(C_{chla})$ ,  $\ln(B_{cyano})$ , and  $\ln(C_{mc})$  variability, respectively (Fig. 2). The best models for each response variable are shown

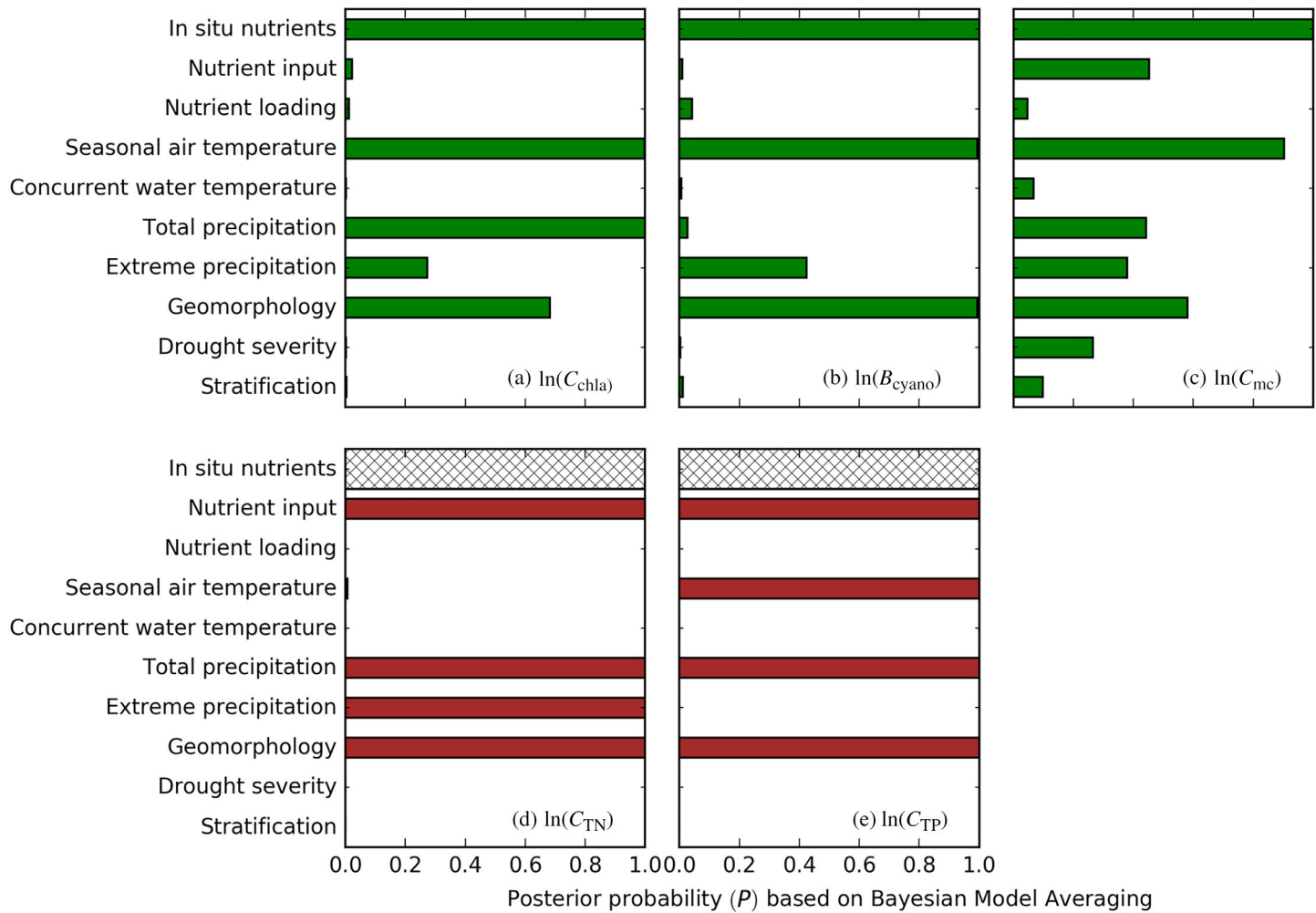
in Table 2. For the best models for  $\ln(C_{chla})$ ,  $\ln(B_{cyano})$ , and  $\ln(C_{mc})$ , the geographic distributions of the contributions of each predictor variable (i.e., each term in the models presented in Table 2) are shown in Figs. 3–5, respectively.

### Variables with highest explanatory power

In situ nutrient concentrations explain the highest amount of variability across HAB metrics considered here (Figs. 3a,b, 4a,b, 5a), and seasonal air temperature (Figs. 3c, 4c, 5b), precipitation (Fig. 3d), and lake geomorphology (Fig. 4d) provide additional explanatory power.

When BMA is used to assess the importance of predictor variables across all possible models, in situ nutrient concentrations and seasonal temperatures are clearly seen as the most robust signals (Fig. 6a–c). Total precipitation and geomorphology are the next two most important factors for  $\ln(C_{chla})$  ( $P = 1.00$  and  $0.68$ , respectively) while geomorphology and extreme precipitation variables are the next most important for  $\ln(B_{cyano})$  ( $P = 1.00$  and  $0.42$ , respectively). For  $\ln(C_{mc})$ , however, BMA is unable to distinguish between the importance of precipitation, geomorphology, and nutrient input variables ( $P = 0.38$ – $0.58$ ). These same variables that are important for explaining variability in HABs are also important for explaining variability in in situ nutrient concentrations (Fig. 6d,e). The variables with the highest probability from BMA for  $\ln(C_{TN})$  and  $\ln(C_{TP})$  are geomorphology and total precipitation, along with nutrient input variables. Notably, seasonal air temperature is also important (tied for the most important category) for the  $\ln(C_{TP})$  model ( $P = 1.00$ ) (Fig. 6e).

While models with seasonal air temperature and precipitation variables have the highest explanatory power (see Table 2;



**Fig. 6.** Posterior probabilities of categories of candidate variables (Yadav et al. 2010) based on BMA (Raftery et al. 1997) of all considered models (see “Methods” section). The posterior probability  $P$  is a measure of the importance of the variables in each category in explaining the observed variability in response variables when accounting for all considered sets of candidate variables. Higher probabilities correspond to greater category importance. See Supporting Information Table S1 for candidate variables considered in each category. The probabilities are computed for each category in explaining each HAB (**a–c**) and nutrient concentration (**d, e**) response variable. Cross-hatching in panels (**d**) and (**e**) indicates categories of candidate variables not considered in model selection for nutrient concentration response variables. Note that these probabilities should not be confused with  $p$  values, which instead denote a level of statistical significance.

Figs. 3–5), and are favored based on BIC and BMA compared to other categories of variables (Fig. 6), the increase in explanatory power over models with other variables is slight when viewed on an absolute basis. For example, the best model for  $\ln(B_{cyano})$  that includes a concurrent water temperature variable has  $R^2 = 0.2788$  whereas the best model in Table 2, which includes spring air temperature, has  $R^2 = 0.2849$ . These relatively small differences in  $R^2$  yield larger differences in BIC, however (for this example, the BIC difference is 10.2). Overall, the temperature and precipitation variables explained 2–7% of the residual variability after accounting for other variables in the model (Supporting Information Fig. S2), and it is the case that the vast majority of models with the lowest BIC scores include seasonal air temperature and/or precipitation variables

(e.g., 92% of models for  $\ln(B_{cyano})$  making up those with cumulative probability  $P = 0.99$  based on BMA). These results provide confidence that the processes represented by seasonal air temperature and precipitation variables are more important for explaining HAB variability.

#### Sign of drift coefficients with best candidate variables

The signs of drift coefficients with seasonal air temperature (Table 2; Figs. 3c, 4c, 5b) indicate that increases in temperature are associated with higher total and cyanobacterial abundance and lower toxicity. Chl  $a$  increased with higher summer air temperature and cyanobacterial biovolume increased with higher spring air temperature, but microcystin in contrast decreased with higher spring air temperature (Table 2, Fig. 5b).

Temperature additionally has a positive association with TP (Table 2, Supporting Information Fig. S4b). The signs of these drift coefficients appear to be robust to other lake and driver conditions, as posterior probabilities from BMA calculated using positive drift coefficient models are all  $P = 1.00$  for Chl *a*, cyanobacterial biovolume, and TP, and the posterior probability calculated using negative drift coefficient models is  $P = 1.00$  for microcystin.

The signs of drift coefficients with precipitation variables indicate that increases in summer extreme precipitation are associated with increases in TN concentrations, while increases in total annual precipitation are associated with increases in Chl *a* concentrations and *decreases* in TN and TP concentrations (Table 2). The positive associations are robust to other lake and driver conditions, as models where summer extreme precipitation and total annual precipitation have positive drift coefficients with TN and Chl *a*, respectively, explain much more of the response variable variability ( $P = 1.00$  from BMA) than negative drift coefficient models. The negative associations between TN and TP and total annual precipitation are similarly robust using the same methodology ( $P = 1.00$ ). We note additionally that the positive association between total annual precipitation and Chl *a* occurs only after accounting for other variables such as TN and TP concentration (Supporting Information Fig. S2); without the influence of other variables, total annual precipitation actually has a negative association with Chl *a* (Supporting Information Fig. S1).

## Discussion

The amount of HAB variability explained by the best models identified here is comparable to, or higher than in, past studies, including those that use only the NLA 2007 data set (Table 3). The improvement in  $R^2$  compared to studies using NLA 2007 data alone is due mainly to the different lake observation data used here, as the same variables used in past studies (i.e., nutrient concentrations) have greater explanatory power for summer months and for 2012 and 2007 together as compared to using all months of 2007 data only. For example,  $\ln(C_{TN})$  has an  $r^2 = 0.23$  with  $\ln(B_{cyano})$  in our study (Supporting Information Fig. S1) compared to  $r^2 = 0.15$  reported elsewhere (Rigosi et al. 2014). As reported in earlier studies, we also find that a higher proportion of the variability in total abundance across lakes (represented by Chl *a*) can be explained using environmental covariates relative to cyanobacterial dominance (represented by cyanobacterial biovolume) or toxicity (represented by microcystin) (Fig. 2). Beyond the variability explained by nutrient concentrations, the strongest explanatory factors of variability in HABs are seasonal air temperature, total and extreme precipitation, and geomorphological variables (Table 2; Figs. 3–5).

Results from model selection (Table 2) are consistent with the hypotheses that higher temperature increases the rates of cyanobacterial growth and internal phosphorus loading and

that longer summers increase the opportunity for cyanobacterial growth (Table 4). Increased summer air temperature is robustly associated with both increased Chl *a* (Fig. 3c, Supporting Information Figs. S1, S2) and TP concentrations (Supporting Information Figs. S1, S2, S4b), while increased spring air temperature is robustly associated with increased cyanobacterial biovolume (Fig. 4c; Supporting Information Figs. S1, S2). The fact that spring air temperature is selected over summer air temperature for the cyanobacterial biovolume model suggests that the impact of longer summers may be more widespread than the impact of increased summertime temperatures on growth rates for influencing cyanobacterial dominance. Longer summers benefit cyanobacterial populations by allowing them to use their buoyancy to establish dominance earlier in the year (Deng et al. 2014). Our findings suggest that this mechanism is of equal or greater importance than the impact on growth rates, in contrast to how these mechanisms have been presented elsewhere (Visser et al. 2016). The association between air temperature and TP concentration also implies that the potential impact of temperature is not primarily in creating conditions conducive to HAB growth, but also in influencing nutrient availability.

Surprisingly, increased spring air temperatures are also associated with *decreased* microcystin concentration (Fig. 5b, Supporting Information Figs. S1, S2), suggesting that longer summers may be increasing cyanobacterial abundance but decreasing harmful toxin concentrations. This is consistent with some observations of how phytoplankton community compositions have changed with longer summers, for instance shifting toward more nontoxic taxa (Anneville et al. 2005). Combined with the reported association between temperature and total abundance, this is also consistent with evidence showing that higher temperatures can lead to faster growth but of less-toxic *Microcystis aeruginosa* cells (Peng et al. 2018). Regardless of the specific mechanism, if this finding can be supported more broadly through subsequent studies, it would have important implications for designing water management efforts, as the harm associated with future blooms influenced by warmer temperatures could be due to higher abundance or dominance of specific species, but not necessarily due to higher toxicity.

The results from model selection are more mixed with regard to the influence of precipitation on HABs (Table 2), as increases in precipitation affect HABs and nutrient concentrations differently depending on the aggregation period of precipitation (summer vs. annual) and the form of precipitation (extreme vs. total). There is evidence that increases in summer extreme precipitation increase nutrient concentrations (Supporting Information Fig. S3d) and that increases in total annual precipitation increase Chl *a* (Fig. 3d, Supporting Information Figs. S1, S2), supporting the load driving hypothesis, while increases in total annual precipitation overall decrease nutrient concentrations (Supporting Information Figs. S3c, S4c), supporting the flushing hypothesis.

Consistent with the load driving hypothesis, higher summer extreme precipitation being robustly associated with

**Table 3.** Summary of selected statistical models in literature using large numbers of lake observations to understand HAB response variables. Each row represents one model, with models organized by whether the response variable represents total abundance, cyanobacterial dominance, or toxicity. The percent of response variable variance explained represents the model  $R^2$  if linear regression is used.  $n$  represents the number of lake observations used to support the model. NLA07 and NLA12 represent the EPA National Lakes Assessment data sets (U.S. Environmental Protection Agency 2010, 2016). EU, SA, and UK indicate samples of European, South American, and United Kingdom lakes, respectively.

Response variable	Percent variance explained	Explanatory variables	$n$	Data used	Model reference
<b>Total abundance</b>					
Chl <i>a</i>	69%	TN, TP, summer air temperature, total annual precipitation	1254	NLA07, NLA12 Jul–Aug	This study
Chl <i>a</i>	41%	TN, TP, surface-water temperature	1156	NLA07	Lin (2017)
Whole-lake annual average remotely sensed Chl <i>a</i>	58%	TN, TP, surface-water temperature	450	2007 Landsat 5	Lin (2017)
Whole-lake annual average remotely sensed Chl <i>a</i>	43%	Maximum daily air temperature, total annual precipitation, average annual precipitation intensity, soil erodibility factor, %cultivated and developed land, shoreline development index, watershed slope, ratio of basin area to lake area	658	2007 Landsat 5	Lin (2017)
Chl <i>a</i>	61%	TP, surface-water temperature, TP $\times$ surface-water temperature	1076	NLA07	Rigosi et al. (2014)
Chl <i>a</i>	59%	TN, surface-water temperature	1142	NLA07	Beaulieu et al. (2013)
Chl <i>a</i>	67%	TN, TP	141	EU, SA lakes	Kosten et al. (2012)
Total biovolume	42%	TN, TP, whole-column water temperature	143	EU, SA lakes	Kosten et al. (2012)
<b>Cyanobacterial dominance</b>					
Cyanobacterial biovolume	28%	TN, TP, spring air temperature, depth, latitude	1208	NLA07, NLA12 Jul–Aug	This study
Cyanobacterial biovolume	15%	TN, surface-water temperature, TN $\times$ surface-water temperature	1076	NLA07	Rigosi et al. (2014)
Cyanobacterial biomass	25%	TN, surface-water temperature	1148	NLA07	Beaulieu et al. (2013)
Cyanobacterial biovolume	34%	TP	807	EU lakes	Carvalho et al. (2013)
Cyanobacterial biovolume	28%	TN, TP, whole-column water temperature	143	EU, SA lakes	Kosten et al. (2012)
Cyanobacterial biovolume	25%	TP, water color, alkalinity, retention time	207	UK lakes	Carvalho et al. (2011)
<b>Toxicity</b>					
Microcystin concentration above $0.05 \mu\text{g L}^{-1}$	25%	TN, spring air temperature	496	NLA07, NLA12 Jul–Aug	This study
Whether microcystin $>0.05 \mu\text{g L}^{-1}$	55%	TN, cyanobacterial biomass, Chl <i>a</i> , dissolved organic carbon, maximum depth, latitude, longitude, % agriculture, % agriculture $\times$ Chl <i>a</i>	1127	NLA07	Taranu et al. (2017)
Microcystin concentration above $0.05 \mu\text{g L}^{-1}$	26%	TN, cyanobacterial biomass, turbidity, lake or reservoir dummy variable, surface-water temperature	362	NLA07	Taranu et al. (2017)

higher TN concentrations would imply that the nutrient runoff resulting in higher nutrient concentrations comes from extreme precipitation events immediately preceding observations. This suggests that summer nutrient concentrations may be driven by precipitation on relatively short timescales and not necessarily by precipitation during spring months, implying that spring loads are either flushed out by the summer or have been taken up by phytoplankton. Extreme precipitation has been identified as an important factor driving TN loading from watersheds previously (Sinha and Michalak 2016; Ballard et al. 2019), and this result confirms the importance of precipitation extremes in driving resultant in-lake nutrient concentrations.

Consistent with the flushing hypothesis, higher total annual precipitation being robustly associated with lower TN and TP concentrations suggests that greater precipitation overall may decrease the residence time of nutrients in lakes and reservoirs, resulting in lower summer TN and TP. This finding expands on the negative relationship between precipitation and Chl *a* observed in Lin (2017), indicating that the flushing effect likely influences the nutrient availability for phytoplankton growth rather than flushing out phytoplankton that have already taken up nutrients. This finding also provides some evidence to suggest that future increases in precipitation may increase flushing and therefore ameliorate bloom conditions in a large number of systems.

With regards to stratification, we did not find evidence for stratification strength or drought severity as widespread environmental drivers of cyanobacterial abundance. None of the stratification or drought severity variables were found to be important either via BMA (Fig. 6) or among the best models selected by BIC (Table 2), for either HABs or nutrients. This is at first surprising given that stratification has previously been identified as an important mechanism in both individual studies (Posch et al. 2012) and more broadly in the literature (Paerl and Huisman 2009). This is also in contrast to a broader synthesis of lakes that found that stratification-related causes explained the spread of cyanobacteria in European peri-alpine lakes (Monchamp et al. 2018). We speculate that the lack of widespread evidence for stratification as a driver could be due to three reasons. First, the discrete depth profiles used to measure stratification could have introduced noise in measuring stratification relative to continuous measurements (e.g., Kraemer et al. 2015), making statistical detection difficult. Second, the diversity of cyanobacterial species in lakes across the continental U.S. could have minimized the observable impacts of stratification, as only some cyanobacterial species can regulate buoyancy and take advantage of stratified conditions (Reynolds 2006). Third, strength of stratification at time of observation may be less important than the duration of stratification periods prior to observation (e.g., Wagner and Adrian 2009). As such, our findings do not necessarily imply that the impact of stratification is less widespread than other mechanisms, but instead underscore the gap in understanding between its impacts in

specific systems and how to measure these impacts relative to other temperature and precipitation mechanisms.

One consistent pattern throughout our results is the importance of variables representing multimonth aggregations in explaining variability in HAB indicators and nutrient concentrations over those based on measurements at single points in time. For example, seasonal air temperature has stronger associations with response variables than concurrent water temperature or stratification (Fig. 6). Because we had access to monthly observations only of air temperature and not water temperature or stratification, the influence of air temperature here is most likely a surrogate for the aggregated impact of water temperature conditions. Evidence for this hypothesis is supported by the fact that shorter aggregations of air temperature (e.g., air temperature during individual months) in general have lower values of  $R^2$  with response variables than multimonth averages of air temperature but higher values of  $R^2$  than water temperature variables. Average spring and summer air temperature could also potentially be indicative of the *duration* of stratification, which has been shown to drive cyanobacterial dominance elsewhere (e.g., Wagner and Adrian 2009; Posch et al. 2012). The importance of multimonth aggregations further explains the positive association between total annual precipitation and Chl *a* only *after* accounting for TN and TP (Supporting Information Fig. S2), as annual precipitation is most likely capturing prior nutrient loading not accounted for by concurrent nutrient concentration measurements. Taken together, the importance of the aggregated environmental driver variables implies a high monitoring burden may be required for efforts seeking to tie HABs to environmental drivers, and suggests a need to identify external data sets with enough temporal coverage to approximate in situ measurements (e.g., data sets interpolated from other in situ measurements like rain gauges, or based on remote sensing).

## Limitations

The findings described here inform the prevalence of hypothesized climatic impacts on HABs across large numbers of lakes, but the approach also has inherent limitations. For instance, the statistical approach used here assumed linear relationships with potential predictors, and was therefore limited in detecting nonlinear relationships such as threshold effects and interaction effects between predictor variables. An examination of regression residuals did not reveal any clear departures from linearity, however (Supporting Information Fig. S2). Moreover, the available observations themselves have limitations, because one summer measurement does not necessarily represent peak bloom intensity nor capture its timing, and observations from the deepest point of the lake may not be representative of phytoplankton bloom conditions more broadly. As well, by only using microcystin observations above the detection limit, this work focused on the processes affecting the severity of toxic blooms (i.e., microcystin concentrations), but not whether or not toxic blooms occur at all (i.e.,



**Table 4.** Hypothesized linkages between climate change and HAB indicators, with expected relationships for candidate variables and resulting evidence from model selection. Evaluated hypotheses are indicated with bold headings (e.g., “Direct growth”) in the first column, and the candidate variables used to test each hypothesis are indicated in the second column (see Supporting Information Table S1 for descriptions of candidate variables). The third column describes the expected relationship, with ↑ representing increase and ↓ representing decrease (e.g., “↑ Predictor ↑ HABs” meaning “an increase in the predictor variable is associated with an increase in HABs”). The fourth column shows the observed result(s) based on the candidate variables selected in model selection and the relationship indicated by the sign of the drift coefficients (see Table 2 caption for descriptions of selected candidate variables).

Climate change impact hypothesis	Candidate variables to test hypothesis	Expected relationship	Observed result(s)
<b>Direct growth</b> Higher temperatures increase cyanobacterial growth rates	$T_{water,surface}; T_{water,bottom};$ $T_{water,mean}; T_{air,JJA}$	↑ Predictor ↑ HABs	↑ $T_{air,JJA}$ ↑ $\ln(C_{chla})$
<b>Longer summer</b> Longer seasons of high temperatures increase the opportunity for cyanobacterial growth	$T_{air,MAM}$	↑ Predictor ↑ HABs	↑ $T_{air,MAM}$ ↑ $\ln(B_{cyano})$
<b>Internal phosphorus loading</b> Higher temperatures increase the rate of internal phosphorus loading in lakes, driving HABs.	$T_{water,surface}; T_{water,bottom};$ $T_{water,mean}; T_{air,JJA}$ $C_{TP}, \ln(C_{TP})$	↑ Predictor ↑ TP ↑ Predictor ↑ HABs	↑ $T_{air,JJA}$ ↑ $\ln(C_{TP})$ ↑ $\ln(C_{TP})$ ↑ $\ln(C_{chla})$ ↑ $\ln(C_{TP})$ ↑ $\ln(B_{cyano})$
<b>Load driving</b> Greater precipitation increases nutrient loading, driving HABs	$P_{annual}; \ln(P_{annual}); P_{MAM};$ $\ln(P_{MAM}); P_{JJA}; \ln(P_{JJA});$ $P_{p>0.90}; P_{p>0.95}; P_{p>0.99};$ $P_{MAM,p>0.90}; P_{MAM,p>0.95};$ $P_{MAM,p>0.99}; P_{JJA,p>0.90}; P_{JJA},$ $p>0.95; P_{JJA,p>0.99}; Q_{TN}; \ln(Q_{TN})$ $C_{TN}; \ln(C_{TN}); C_{TP}; \ln(C_{TP})$	↑ Predictor ↑ TN or ↑ Predictor ↑ TP ↑ Predictor ↑ HABs	↑ $P_{JJA,p>0.95}$ ↑ $\ln(C_{TN})$ ↑ $\ln(C_{TN})$ ↑ $\ln(C_{chla})$ ↑ $\ln(C_{TP})$ ↑ $\ln(C_{chla})$ ↑ $\ln(C_{TN})$ ↑ $\ln(B_{cyano})$ ↑ $\ln(C_{TP})$ ↑ $\ln(B_{cyano})$ ↑ $\ln(C_{TN})$ ↑ $\ln(C_{mc})$
<b>Flushing</b> Greater precipitation decreases HABs by reducing residence times	$P_{annual}; \ln(P_{annual}); P_{MAM};$ $\ln(P_{MAM}); P_{JJA}; \ln(P_{JJA}); P_{p>0.90};$ $P_{p>0.95}; P_{p>0.99}; P_{MAM,p>0.90};$ $P_{MAM,p>0.95}; P_{MAM,p>0.99};$ $P_{JJA,p>0.90}; P_{JJA,p>0.95}; P_{JJA},$ $p>0.99; Q_{TN}; \ln(Q_{TN})$ $C_{TN}; \ln(C_{TN}); C_{TP}; \ln(C_{TP})$	↑ Predictor ↓ TN or ↑ Predictor ↓ TP ↑ Predictor ↑ HABs	↑ $P_{annual}$ ↓ $\ln(C_{TN})$ ↑ $\ln(P_{annual})$ ↓ $\ln(C_{TP})$ ↑ $\ln(C_{TN})$ ↑ $\ln(C_{chla})$ ↑ $\ln(C_{TP})$ ↑ $\ln(C_{chla})$ ↑ $\ln(C_{TN})$ ↑ $\ln(B_{cyano})$ ↑ $\ln(C_{TP})$ ↑ $\ln(B_{cyano})$ ↑ $\ln(C_{TN})$ ↑ $\ln(C_{mc})$
<b>Stratification</b> Greater vertical stratification selects for cyanobacteria	$D_{thermocline}; \ln(D_{thermocline});$ $f_{buoyancy}; \Delta T_{surface-bottom}$	↑ Predictor ↑ HABs	Listed candidate variables are not included in the best models determined by BIC

microcystin detectability). In addition, environmental drivers are also not fully characterized; for example, there is little information available to estimate lake mixing, and also limited measurements of stratification. These limitations demonstrate that there are still substantial gaps in understanding of underlying factors driving bloom conditions in large numbers of lakes.

Despite these limitations, the study's broad geographic scope provides support for specific climate impact hypotheses, and this can provide guidance to mechanistic modelers and water managers for differentiating between various climate impact hypotheses. While statistical approaches such as the one used here have the advantage of providing more reproducible results compared to mechanistic modeling approaches and also of allowing for easier revision as data sets improve (Anderson et al. 2015), they are limited in investigating more complex interactions between different phytoplankton species and stressors as compared to coupled physical-biological models (Jöhnk et al. 2008). However, given that full mechanistic models simulating phytoplankton population dynamics or toxin production are still rare (Anderson et al. 2015), the findings described here contribute to the understanding of dominant processes underlying HAB formation that can then inform future implementation of mechanistic models.

## Conclusions

Overall, our findings emphasize the importance of nutrient input reductions for combatting the impact of climate change on HABs. While we found support for some mechanisms indicating that changes in future climate may ameliorate bloom conditions (e.g., flushing potentially reducing nutrient concentrations and warmer temperatures potentially reducing toxicity), the majority of mechanisms identified here are expected to exacerbate HAB conditions. At the same time, nutrient concentrations explain the most variability in HABs, and nutrient inputs (e.g., fertilizer) are robustly associated with nutrient concentrations. Reducing nutrient inputs therefore has the potential to reduce sensitivity to future changes in climate, both by reducing the nutrients available to precipitation for generating runoff and by reducing the reservoir of legacy nutrients for temperature-mediated internal loading.

Support for the internal phosphorus loading hypothesis further suggests that the paradigm considering nutrient availability and temperature as separate factors (Brookes and Carey 2011; Kosten et al. 2012; Rigosi et al. 2014) may be incomplete. The linkage between temperature and nutrient concentration described in this work provides new context for past efforts examining the importance of temperature vs. nutrient availability, and implies that the effect of temperature on nutrient availability also needs to be taken into consideration. As the importance of internal phosphorus loading is likely to increase with rising temperatures, this moreover indicates a need to understand the role of internal phosphorus loading in lake nutrient budgets. For instance, in Lake Taihu, internal

soluble reactive phosphorus loading has substantially changed TN/TP ratios, with internal phosphorus loading accounting for 54% of the increase in TP observed in the water column (Ding et al. 2018). Internal phosphorus loading is especially important to consider in the context of assessing external nutrient reductions. Even if nutrient inputs were completely eliminated immediately, the impact of warming could continue to stimulate internal phosphorus loading (Jeppesen et al. 2007), potentially impacting the timescale of expected system recovery (Ho and Michalak 2017). The ongoing effect of nutrient legacies in watersheds (Sharpley et al. 2014) and their contribution to internal nutrient loading means it is even more urgent to reduce nutrient inputs broadly.

## References

- Abatzoglou, J. T. 2013. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* **33**: 121–131. doi:[10.1002/joc.3413](https://doi.org/10.1002/joc.3413)
- Abatzoglou, J. T., R. Barbero, J. W. Wolf, and Z. A. Holden. 2014. Tracking interannual streamflow variability with drought indices in the U.S. Pacific Northwest. *J. Hydrometeorol.* **15**: 1900–1912. doi:[10.1175/JHM-D-13-0167.1](https://doi.org/10.1175/JHM-D-13-0167.1)
- Anderson, C. R., S. K. Moore, M. C. Tomlinson, J. Silke, and C. K. Cusack. 2015. Living with harmful algal blooms in a changing world: Strategies for modeling and mitigating their effects in coastal marine ecosystems, p. 495–561. *In* J. F. Shroder, J. T. Ellis, and D. J. Sherman [eds.], *Coastal and marine hazards, risks, and disasters*. Elsevier.
- Anderson, D. M., P. M. Glibert, and J. M. Burkholder. 2002. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries* **25**: 704–726. doi:[10.1007/BF02804901](https://doi.org/10.1007/BF02804901)
- Anneville, O., S. Gammeter, and D. Straille. 2005. Phosphorus decrease and climate variability: Mediators of synchrony in phytoplankton changes among European peri-alpine lakes. *Freshw. Biol.* **50**: 1731–1746. doi:[10.1111/j.1365-2427.2005.01429.x](https://doi.org/10.1111/j.1365-2427.2005.01429.x)
- Ballard, T. C., E. Sinha, and A. M. Michalak. 2019. Long-term changes in precipitation and temperature have already impacted nitrogen loading. *Environ. Sci. Technol.* **53**: 5080–5090. doi:[10.1021/acs.est.8b06898](https://doi.org/10.1021/acs.est.8b06898)
- Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. *Limnol. Oceanogr.* **58**: 1736–1746. doi:[10.4319/lo.2013.58.5.1736](https://doi.org/10.4319/lo.2013.58.5.1736)
- Brookes, J. D., and C. C. Carey. 2011. Resilience to blooms. *Science* **334**: 46–47. doi:[10.1126/science.1207349](https://doi.org/10.1126/science.1207349)
- Cantin, A., B. E. Beisner, J. M. Gunn, Y. T. Prairie, and J. G. Winter. 2011. Effects of thermocline deepening on lake plankton communities. *Can. J. Fish. Aquat. Sci.* **68**: 260–276. doi:[10.1139/F10-138](https://doi.org/10.1139/F10-138)
- Carey, C. C., B. W. Ibelings, E. P. Hoffmann, D. P. Hamilton, and J. D. Brookes. 2012. Eco-physiological adaptations that

- favour freshwater cyanobacteria in a changing climate. *Water Res.* **46**: 1394–1407. doi:[10.1016/j.watres.2011.12.016](https://doi.org/10.1016/j.watres.2011.12.016)
- Carvalho, L., C. A. Miller, E. M. Scott, G. A. Codd, P. S. Davies, and A. N. Tyler. 2011. Cyanobacterial blooms: Statistical models describing risk factors for national-scale lake assessment and lake management. *Sci. Total Environ.* **409**: 5353–5358. doi:[10.1016/j.scitotenv.2011.09.030](https://doi.org/10.1016/j.scitotenv.2011.09.030)
- Carvalho, L., C. McDonald, C. de Hoyos, and others. 2013. Sustaining recreational quality of European lakes: Minimizing the health risks from algal blooms through phosphorus control. *J. Appl. Ecol.* **50**: 315–323. doi:[10.1111/1365-2664.12059](https://doi.org/10.1111/1365-2664.12059)
- Chapra, S. C., and others. 2017. Climate change impacts on harmful algal blooms in U.S. freshwaters: A screening-level assessment. *Environ. Sci. Technol.* **51**: 8933–8943. doi:[10.1021/acs.est.7b01498](https://doi.org/10.1021/acs.est.7b01498)
- Chorus, I., and J. Bartram. 1999. Toxic cyanobacter in water: A guide to their public health consequences, monitoring and management. E & FN Spon.
- Del Giudice, D., Y. Zhou, E. Sinha, and A. M. Michalak. 2018. Long-term phosphorus loading and springtime temperatures explain interannual variability of hypoxia in a large temperate lake. *Environ. Sci. Technol.* **52**: 2046–2054. doi:[10.1021/acs.est.7b04730](https://doi.org/10.1021/acs.est.7b04730)
- Deng, J., B. Qin, H. W. Paerl, Y. Zhang, J. Ma, and Y. Chen. 2014. Earlier and warmer springs increase cyanobacterial (*Microcystis* spp.) blooms in subtropical Lake Taihu, China. *Freshw. Biol.* **59**: 1076–1085. doi:[10.1111/fwb.12330](https://doi.org/10.1111/fwb.12330)
- Descy, J. P., and others. 2016. Identifying the factors determining blooms of cyanobacteria in a set of shallow lakes. *Ecol. Inform.* **34**: 129–138. doi:[10.1016/j.ecoinf.2016.05.003](https://doi.org/10.1016/j.ecoinf.2016.05.003)
- Ding, S., and others. 2018. Internal phosphorus loading from sediments causes seasonal nitrogen limitation for harmful algal blooms. *Sci. Total Environ.* **625**: 872–884. doi:[10.1016/j.scitotenv.2017.12.348](https://doi.org/10.1016/j.scitotenv.2017.12.348)
- Fry, J., and others. 2011. Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sensing* **77**: 858–864.
- Geh, E. N., A. A. De La Cruz, D. Ghosh, G. Stelma, and J. A. Bernstein. 2016. Sensitization of a child to Cyanobacteria after recreational swimming in a lake. *J. Allergy Clin. Immunol.* **137**: 1902–1904. doi:[10.1016/j.jaci.2015.12.1340](https://doi.org/10.1016/j.jaci.2015.12.1340)
- Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**: 18–27. doi:[10.1016/j.rse.2017.06.031](https://doi.org/10.1016/j.rse.2017.06.031)
- Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* **84**: 2809–2815. doi:[10.1890/02-3114](https://doi.org/10.1890/02-3114)
- Hayes, N. M., M. J. Vanni, M. J. Horgan, and W. H. Renwick. 2015. Climate and land use interactively affect lake phytoplankton nutrient limitation status. *Ecology* **96**: 392–402. doi:[10.1890/13-1840.1](https://doi.org/10.1890/13-1840.1)
- Ho, J. C., and A. M. Michalak. 2017. Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. *J. Great Lakes Res.* **43**: 221–228. doi:[10.1016/j.jglr.2017.04.001](https://doi.org/10.1016/j.jglr.2017.04.001)
- Homer, C. G., and others. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information. *Photogramm. Eng. Remote Sensing* **81**: 345–354. doi:[10.14358/PERS.81.5.345](https://doi.org/10.14358/PERS.81.5.345)
- Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* **16**: 471–483. doi:[10.1038/s41579-018-0040-1](https://doi.org/10.1038/s41579-018-0040-1)
- Jeppesen, E., M. Søndergaard, M. Meerhoff, T. L. Lauridsen, and J. P. Jensen. 2007. Shallow lake restoration by nutrient loading reduction—some recent findings and challenges ahead. *Hydrobiologia* **584**: 239–252. doi:[10.1007/s10750-007-0596-7](https://doi.org/10.1007/s10750-007-0596-7)
- Jöhnk, K. D., J. Huisman, J. Sharples, B. Sommeijer, P. M. Visser, and J. M. Stroom. 2008. Summer heatwaves promote blooms of harmful cyanobacteria. *Glob. Chang. Biol.* **14**: 495–512. doi:[10.1111/j.1365-2486.2007.01510.x](https://doi.org/10.1111/j.1365-2486.2007.01510.x)
- Kosten, S., and others. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob. Chang. Biol.* **18**: 118–126. doi:[10.1111/j.1365-2486.2011.02488.x](https://doi.org/10.1111/j.1365-2486.2011.02488.x)
- Kraemer, B. M., and others. 2015. Morphometry and average temperature affect lake stratification responses to climate change. *Geophys. Res. Lett.* **42**: 4981–4988. doi:[10.1002/2015GL064097](https://doi.org/10.1002/2015GL064097)
- Kudela, R. M., and others. 2015. Harmful algal blooms. A scientific summary for policy makers. IOC/UNESCO.
- Leech, D. M., A. I. Pollard, S. G. Labou, and S. E. Hampton. 2018. Fewer blue lakes and more murky lakes across the continental U.S.: Implications for planktonic food webs. *Limnol. Oceanogr.* **63**: 2661–2680. doi:[10.1002/LNO.10967](https://doi.org/10.1002/LNO.10967)
- Lin, S. 2017. Climate change and algal blooms. Ph.D. thesis. Michigan State Univ.
- Mantzouki, E., and others. 2018. Temperature effects explain continental scale distribution of cyanobacterial toxins. *Toxins* **10**: 1–24. doi:[10.3390/toxins10040156](https://doi.org/10.3390/toxins10040156)
- Marion, J. W., F. Zhang, D. Cutting, and J. Lee. 2017. Associations between county-level land cover classes and cyanobacteria blooms in the United States. *Ecol. Eng.* **108**: 556–563. doi:[10.1016/j.ecoleng.2017.07.032](https://doi.org/10.1016/j.ecoleng.2017.07.032)
- Matisoff, G., and others. 2016. Internal loading of phosphorus in western Lake Erie. *J. Great Lakes Res.* **42**: 775–788. doi:[10.1016/j.jglr.2016.04.004](https://doi.org/10.1016/j.jglr.2016.04.004)
- Michalak, A. M. 2016. Study role of climate change in extreme threats to water quality. *Nature* **535**: 349–350. doi:[10.1038/535349a](https://doi.org/10.1038/535349a)
- Michalak, A. M., and others. 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc. Natl. Acad. Sci. USA* **110**: 6448–6452. doi:[10.1073/pnas.1216006110](https://doi.org/10.1073/pnas.1216006110)

- Monchamp, M. E., P. Spaak, I. Domaizon, N. Dubois, D. Bouffard, and F. Pomati. 2018. Homogenization of lake cyanobacterial communities over a century of climate change and eutrophication. *Nat. Ecol. Evol.* **2**: 317–324. doi:[10.1038/s41559-017-0407-0](https://doi.org/10.1038/s41559-017-0407-0)
- Motew, M., E. G. Booth, S. R. Carpenter, X. Chen, and C. J. Kucharik. 2018. The synergistic effect of manure supply and extreme precipitation on surface water quality. *Environ. Res. Lett.* **13**: 044016. doi:[10.1088/1748-9326/aaade6](https://doi.org/10.1088/1748-9326/aaade6)
- O'Reilly, C. M., and others. 2015. Rapid and highly variable warming of lake surface waters around the globe. *Geophys. Res. Lett.* **42**: 10773–10781. doi:[10.1002/2015GL066235](https://doi.org/10.1002/2015GL066235)
- Orihel, D. M., H. M. Baulch, N. J. Casson, R. L. North, C. T. Parsons, D. C. M. Seckar, and J. J. Venkiteswaran. 2017. Internal phosphorus loading in Canadian fresh waters: A critical review and data analysis. *Can. J. Fish. Aquat. Sci.* **74**: 2005–2029. doi:[10.1139/cjfas-2016-0500](https://doi.org/10.1139/cjfas-2016-0500)
- Paerl, H. W., and J. Huisman. 2008. Blooms like it hot. *Science* **320**: 57–58. doi:[10.1126/science.1155398](https://doi.org/10.1126/science.1155398)
- Paerl, H. W., and J. Huisman. 2009. Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environ. Microbiol. Rep.* **1**: 27–37. doi:[10.1111/j.1758-2229.2008.00004.x](https://doi.org/10.1111/j.1758-2229.2008.00004.x)
- Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. *Water Res.* **46**: 1349–1363. doi:[10.1016/j.watres.2011.08.002](https://doi.org/10.1016/j.watres.2011.08.002)
- Paerl, H. W., W. S. Gardner, K. E. Havens, A. R. Joyner, M. J. McCarthy, S. E. Newell, B. Qin, and J. T. Scott. 2016. Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. *Harmful Algae* **54**: 213–222. doi:[10.1016/j.hal.2015.09.009](https://doi.org/10.1016/j.hal.2015.09.009)
- Peck, D. V., A. R. Olsen, M. H. Weber, S. G. Paulsen, C. Peterson, and S. M. Holdsworth. 2013. Survey design and extent estimates for the National Lakes Assessment. *Freshw. Sci.* **32**: 1231–1245. doi:[10.1899/11-075.1](https://doi.org/10.1899/11-075.1)
- Peng, G., R. M. Martin, S. P. Dearth, X. Sun, G. L. Boyer, S. R. Campagna, S. Lin, and S. W. Wilhelm. 2018. Seasonally relevant cool temperatures interact with N chemistry to increase microcystins produced in lab cultures of *Microcystis aeruginosa* NIES-843. *Environ. Sci. Technol.* **52**: 4127–4136. doi:[10.1021/acs.est.7b06532](https://doi.org/10.1021/acs.est.7b06532)
- Posch, T., O. Köster, M. M. Salcher, and J. Pernthaler. 2012. Harmful filamentous cyanobacteria favoured by reduced water turnover with lake warming. *Nat. Clim. Chang.* **2**: 809–813. doi:[10.1038/nclimate1581](https://doi.org/10.1038/nclimate1581)
- Raftery, A. E. 1995. Bayesian model selection in social research. *Sociol. Methodol.* **25**: 111–163. doi:[10.2307/271063](https://doi.org/10.2307/271063)
- Raftery, A. E., D. Madigan, and J. A. Hoeting. 1997. Bayesian model averaging for linear regression models. *J. Am. Stat. Assoc.* **92**: 179–191. doi:[10.1080/01621459.1997.10473615](https://doi.org/10.1080/01621459.1997.10473615)
- Read, J. S., D. P. Hamilton, I. D. Jones, K. Muraoka, L. A. Winslow, R. Kroiss, C. H. Wu, and E. Gaiser. 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environ. Model. Softw.* **26**: 1325–1336. doi:[10.1016/j.envsoft.2011.05.006](https://doi.org/10.1016/j.envsoft.2011.05.006)
- Reichwaldt, E. S., and A. Ghadouani. 2012. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: Between simplistic scenarios and complex dynamics. *Water Res.* **46**: 1372–1393. doi:[10.1016/j.watres.2011.11.052](https://doi.org/10.1016/j.watres.2011.11.052)
- Reynolds, C. S. 2006. The ecology of phytoplankton. Cambridge Univ. Press.
- Rigosi, A., C. C. Carey, B. W. Ibelings, and J. D. Brookes. 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. *Limnol. Oceanogr.* **59**: 99–114. doi:[10.4319/lo.2014.59.01.0099](https://doi.org/10.4319/lo.2014.59.01.0099)
- Rigosi, A., and others. 2015. Determining the probability of cyanobacterial blooms: The application of Bayesian networks in multiple lake systems. *Ecol. Appl.* **25**: 186–199. doi:[10.1890/13-1677.1](https://doi.org/10.1890/13-1677.1)
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* **6**: 461–464. doi:[10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
- Sharpley, A., H. P. Jarvie, A. Buda, L. May, B. Spears, and P. Kleinman. 2014. Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. *J. Environ. Qual.* **42**: 1308–1326. doi:[10.2134/jeq2013.03.0098](https://doi.org/10.2134/jeq2013.03.0098)
- Sinha, E., and A. M. Michalak. 2016. Precipitation dominates interannual variability of riverine nitrogen loading across the continental United States. *Environ. Sci. Technol.* **50**: 12874–12884. doi:[10.1021/acs.est.6b04455](https://doi.org/10.1021/acs.est.6b04455)
- Swaney, D. P., B. Hong, and R. W. Howarth. 2018a. NANI/NAPI Calculator Toolbox version 3.1 documentation.
- Swaney, D. P., R. W. Howarth, and B. Hong. 2018b. County, subregional and regional nitrogen data derived from the Net Anthropogenic Nitrogen Inputs (NANI) toolbox. *Data Brief* **18**: 1877–1888. doi:[10.1016/j.dib.2018.04.098](https://doi.org/10.1016/j.dib.2018.04.098)
- Taranu, Z. E., R. W. Zurawell, F. Pick, and I. Gregory-Eaves. 2012. Predicting cyanobacterial dynamics in the face of global change: The importance of scale and environmental context. *Glob. Chang. Biol.* **18**: 3477–3490. doi:[10.1111/gcb.12015](https://doi.org/10.1111/gcb.12015)
- Taranu, Z. E., and others. 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. *Ecol. Lett.* **18**: 375–384. doi:[10.1111/ele.12420](https://doi.org/10.1111/ele.12420)
- Taranu, Z. E., I. Gregory-Eaves, R. J. Steele, M. Beaulieu, and P. Legendre. 2017. Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework for modelling an important health risk factor. *Glob. Ecol. Biogeogr.* **26**: 625–637. doi:[10.1111/geb.12569](https://doi.org/10.1111/geb.12569)
- Thornton, P. E., M. M. Thornton, B. W. Mayer, Y. Wei, R. Devarakonda, R. S. Vose, and R. B. Cook. 2017. Daymet: Daily surface weather data on a 1-km grid for North America, version 3, ORNL DAAC.
- U.S. Environmental Protection Agency. 2007. Survey of the nation's lakes. Field operations manual. EPA 841-B-07-004.

- <https://www.epa.gov/national-aquatic-resource-surveys/manuals-used-national-aquatic-resource-surveys#National%20Lakes%20Assessment>
- U.S. Environmental Protection Agency. 2010. National lakes assessment 2007 (data and metadata files). National Aquatic Resource Surveys.
- U.S. Environmental Protection Agency. 2011. 2012 national lakes assessment. Field operations manual. EPA-841-B-11-003.
- U.S. Environmental Protection Agency. 2016. National lakes assessment 2012 (data and metadata files). National Aquatic Resource Surveys.
- Visser, P. M., J. M. H. Verspagen, G. Sandrini, L. J. Stal, H. C. P. Matthijs, T. W. Davis, H. W. Paerl, and J. Huisman. 2016. How rising CO<sub>2</sub> and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae* **54**: 145–159. doi:[10.1016/j.hal.2015.12.006](https://doi.org/10.1016/j.hal.2015.12.006)
- Wagner, C., and R. Adrian. 2009. Cyanobacteria dominance: Quantifying the effects of climate change. *Limnol. Oceanogr.* **54**: 2460–2468. doi:[10.4319/lo.2009.54.6\\_part\\_2.2460](https://doi.org/10.4319/lo.2009.54.6_part_2.2460)
- Winder, M. 2012. Lake warming mimics fertilization. *Nat. Clim. Chang.* **2**: 771–772. doi:[10.1038/nclimate1728](https://doi.org/10.1038/nclimate1728)
- Xia, Y., and others. 2009. NLDAS primary forcing data L4 hourly 0.125 x 0.125 degree V002. NASA/GSFC/HSL Goddard Earth Sciences Data and Information Services Center (GES DISC).
- Xia, Y., and others. 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. Atmos.* **117**: D03109. doi:[10.1029/2011JD016048](https://doi.org/10.1029/2011JD016048)
- Yadav, V., K. L. Mueller, D. Dragoni, and A. M. Michalak. 2010. A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America. *Biogeosciences* **7**: 2655–2671. doi:[10.5194/bg-7-2655-2010](https://doi.org/10.5194/bg-7-2655-2010)
- Zhou, Y., A. M. Michalak, D. Beletsky, Y. R. Rao, and R. P. Richards. 2015. Record-breaking Lake Erie hypoxia during 2012 drought. *Environ. Sci. Technol.* **49**: 800–807. doi:[10.1021/es503981n](https://doi.org/10.1021/es503981n)

#### Acknowledgments

We thank E. Sinha and T. Ballard for discussions and helpful input, and E. Read and J. Hollister for answering question about NLA data. The National Lakes Assessment 2007 and 2012 data were the result of the collective efforts of dedicated field crews, laboratory staff, data management and quality control staff, analysts, and many others from EPA, states, tribes, federal agencies, universities, and other organizations. Please contact [nars-hq@epa.gov](mailto:nars-hq@epa.gov) with any questions. The temperature data used in this effort were acquired as part of the activities of NASA's Science Mission Directorate, and are archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

#### Conflict of Interest

None declared.

*Submitted 18 September 2018*

*Revised 23 March 2019*

*Accepted 02 October 2019*

*Associate editor: John Downing*