Water Research 229 (2023) 119402

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Large scale seasonal forecasting of peak season algae metrics in the Midwest and Northeast U.S.

Maxwell R.W. Beal [a,*], Grace M. Wilkinson [b], Paul J. Block [a]

[a] Department of Civil and Environmental Engineering, University of Wisconsin – Madison, 1415, Engineering Dr., Madison, WI 53706, United States
[b] Center for Limnology, University of Wisconsin – Madison, 680N Park St, Madison, WI 53706, United States

ARTICLE INFO

Keywords:
Season-ahead forecasting
Algae
Inland lakes
Hydroclimate
LAGOS-NE

ABSTRACT

In recent decades, many inland lakes have seen an increase in the prevalence of potentially harmful algae. In many inland lakes, the peak season for algae abundance (summer and early fall in the northern hemisphere) coincides with the peak season for recreational use. Currently, little information regarding expected algae conditions is available prior to the peak season for productivity in inland lakes. Peak season algae conditions are influenced by an array of pre-season (spring and early summer) local and global scale variables; identifying these variables for forecast development may be useful in managing potential public health threats posed by harmful algae. Using the LAGOS-NE dataset, pre-season local and global drivers of peak-season algae metrics (represented by chlorophyll-a) are identified for 178 lakes across the Northeast and Midwest U.S. from readily available gridded datasets. Forecasting models are built for each lake conditioned on relevant pre-season predictors. Forecasts are assessed for the *magnitude, severity*, and *duration* of seasonal chlorophyll concentrations. Regions of pre-season sea surface temperature, and pre-season chlorophyll-a demonstrate the most predictive power for peak season algae metrics, and resulting models show significant skill. Based on categorical forecast metrics, more than 70% of magnitude models and 90% of duration models outperform climatology. Forecasts of high and severe algae magnitude perform best in large mesotrophic and oligotrophic lakes, however, high algae duration performance appears less dependent on lake characteristics. The advance notice of elevated algae biomass provided by these models may allow lake managers to better prepare for challenges posed by algae during the high use season for inland lakes.

1. Introduction

Rapid proliferation of algae in surface freshwaters has negative consequences for ecosystem function (Sunda et al., 2006; Huisman et al., 2018), economic opportunity (Dodds et al., 2009), and human health due to the potential for toxin production in some species (Carmichael, 2001; Carmichael and Boyer, 2016). In recent decades, anthropogenic disturbance of nitrogen and phosphorus cycles has resulted in widespread eutrophication, leading to an increase in the prevalence of harmful algae (Smith, 2003; O'Neil et al., 2012; Paerl and Paul, 2012). For many waterbodies, hydroclimatic variability plays an important role in determining water quality on inter- and intra-annual timescales, and may influence the suitability of conditions for algae growth (León-Muñoz et al., 2018; Scordo et al., 2022). Nutrient runoff, in particular, is sensitive to variability in the hydrologic cycle, which has been projected to intensify with climate change (Glavan et al., 2015; Me et al., 2018). Anthropogenic stressors favoring the dominance of harmful algae, combined with notable variability in algae biomass, presents a substantial challenge for water resource managers. In the U.S., harmful algae in large waterbodies such as Lake Erie has received significant research and media attention (Reutter et al., 2011; International Joint Commission, 2014; Wines, 2014; Patel and Parshina-Kottas, 2017; Dalton, 2021), however, despite similar concerns, strategies for managing harmful algae in small inland waterbodies across the U.S. have received less attention (Brooks et al., 2016).

In the northern hemisphere, algae biomass tends to peak in the late summer and early fall (July-October) as a result of a complex array of pre-season and within-season physical, chemical, and biological processes. In the Midwest and Northeast U.S., this season is characterized by warm temperatures and increased sunlight, allowing for increased photosynthesis and algae productivity (Singh and Singh, 2015). In many instances, significant intra- and inter-annual variability in peak season

* Corresponding author.
E-mail address: mrbeal@wisc.edu (M.R.W. Beal).

https://doi.org/10.1016/j.watres.2022.119402
Received 1 September 2022; Received in revised form 16 November 2022; Accepted 20 November 2022
Available online 22 November 2022
0043-1354/

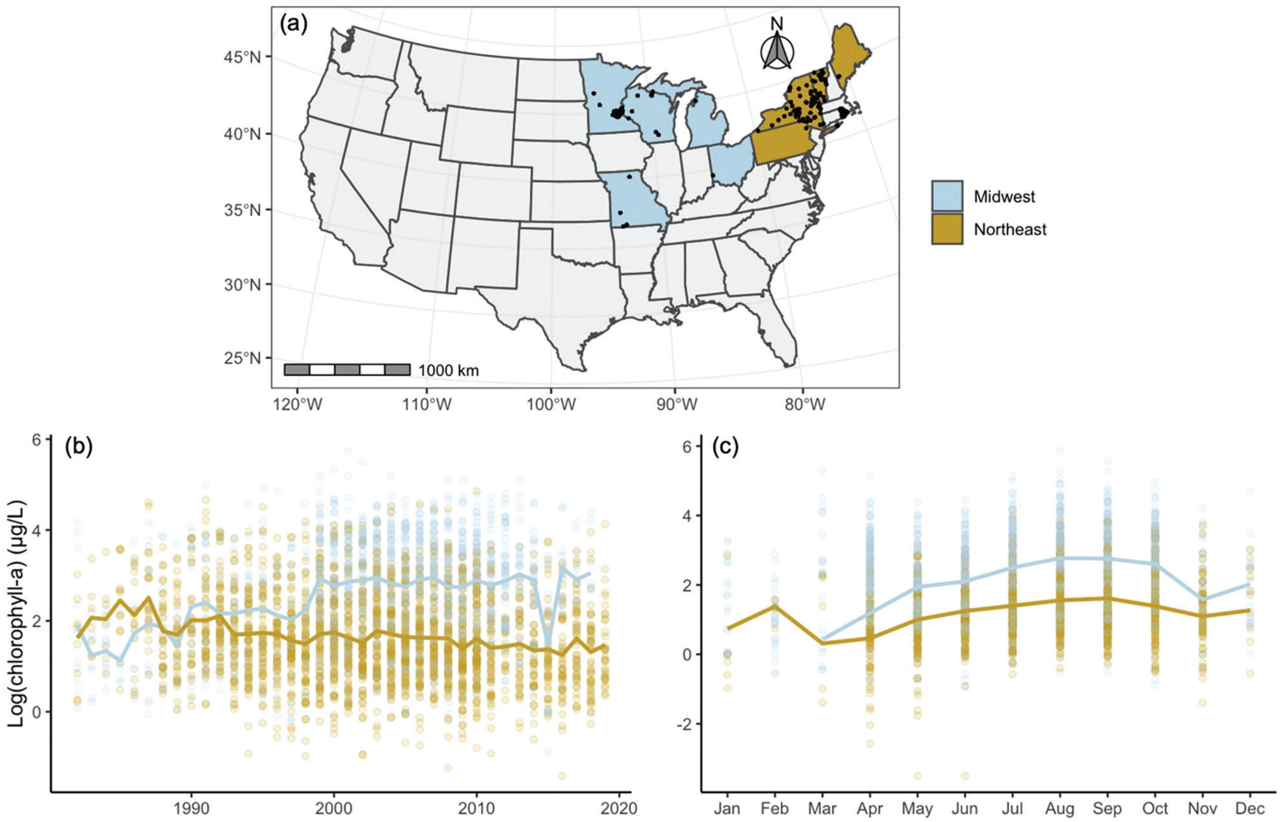

Fig. 1. Candidate lakes for model development in the Northeast and Midwest U.S. (a), including mean JASO chlorophyll-a (µg/L) concentrations (b) and mean monthly chlorophyll-a values (c). Points in (b) and (c) represent values for each lake.

algae biomass is evident, driven partially by local hydrology and temperature that are in turn modulated by large scale climate phenomena through atmospheric teleconnections (Beal et al., 2021). Predictions of how these algae conditions vary may benefit lake managers by allowing them to take early actions to reduce or mitigate harm caused by intense algae growth. Short-term (days to weeks) predictions of chlorophyll-*a* (a proxy for algal biomass) are typically issued within-season and focus on expected bloom formation or toxin production, allowing managers to take rapid actions to address odor and taste issues, transition to alternative water sources, post warning signs at beaches, etc. (Zhang et al., 2013; Chen et al., 2015; Qian et al., 2021; Wynne et al., 2013). In contrast, longer-lead (months) pre-season predictions of expected algae conditions may allow lake managers to address a different set of actions (e.g. life-guard training, public awareness, etc.) and decisions, (e.g., testing and monitoring budgets and plans). Together, these predictions can provide decision makers with multi-scale information to inform appropriate actions at various lead times. However, season-ahead predictions for water quality have received relatively little attention.

Longer-lead predictions of oceanic chlorophyll-*a* and inland nutrient loading have been developed with some success (e.g. Cho et al., 2016; Park et al., 2019; Rousseaux et al., 2021), but little attention has been devoted to inland lakes. Long-lead predictions of algae that do exist typically focus on singular metrics (often mean biomass) to characterize the potential loss of ecosystem services due to algae accumulation, however, further characterization may be warranted. For example, Wilkinson et al. (2021) define three metrics to characterize algae conditions, including: *magnitude* (mean seasonal chlorophyll), *severity* (peak seasonal chlorophyll), and *duration* (length of time chlorophyll is above a threshold concentration). In addition to information provided by mean biomass, this approach provides lake managers with information that specifically addresses two key management concerns related to algae: the potential for severe consequences of algae blooms such as fish kills and toxin production, and length of time a lake may be unfit for recreation. Long-lead predictions also often rely predominantly on nutrient loads as predictors (Stow et al., 1997; Lathrop et al., 1998; Stumpf et al., 2016), however consideration of relevant hydroclimatic predictors has the potential to enhance prediction performance and expand the availability of water quality predictions to many small inland lakes (Beal et al., 2021). There is a large and long body of evidence illustrating the impacts of external nutrient loading on phytoplankton growth in lakes (Vollenweider, 1971; Schindler, 1978; Reynolds, 1984; Elliott et al., 2006; Kane et al., 2014). Phosphorus and nitrogen are widely considered the most important nutrients for phytoplankton growth in freshwater (Schindler, 1971, 1977). Transportation of phosphorus and nitrogen into a lake from the surrounding watershed is an important driver of algae abundance in many systems. Nutrient transport is influenced by global and local hydroclimatic variables, and thus also represent important processes in determining algae abundance. Increased precipitation has been linked to increased fluxes of nitrogen and phosphorus (Sinha et al., 2017), particularly in extreme precipitation events (Haygarth and Jarvis, 1997; Royer et al., 2006; Carpenter et al., 2015, 2018). Soil moisture conditions may also influence nutrient loading by regulating runoff potential (Kleinman et al., 2006; Liu et al., 2014). Finally, water temperature has also been shown to control phytoplankton biomass and growth rate (Eppley, 1972; Konopka and Brock, 1978; Robarts and Zohary, 1987; Elliott et al., 2006; Liu et al., 2019; Trombetta et al., 2021), and is closely linked to local air temperature (Shuter et al., 1983; Woolway et al., 2020; Zhu et al., 2020). March-June water temperature data are not readily available for all study lakes and are therefore not included in the final set of potential predictors.

In addition to management applications, season-ahead forecasts at scale provide a unique opportunity to understand ecological relationships between hydroclimatic variables and water quality (Houlahan et al., 2017). In particular, the relevance of global scale processes in determining algae biomass in inland lakes is not well studied. Several studies have identified teleconnections between large-scale climate phenomena and phytoplankton dynamics in inland lakes (Arhonditsis et al., 2004; Xiao et al., 2019; da Rosa Wieliczko et al., 2021), however, few studies exist that investigate the application of global climate patterns to chlorophyll-*a* prediction in inland lakes. Beal et al. (2021), developed a sub-seasonal (2-month lead) forecast of cyanobacteria biomass in Lake Mendota, Wisconsin (WI), conditioned on local hydroclimatic variables and teleconnections with global climate patterns. A large-scale analysis of season-ahead predictors of algae biomass is well suited to improve the understanding of dominant climate signals related to chlorophyll-*a*.

Using chlorophyll-a time series from 178 lakes in the Northeast and Midwest U.S. we evaluated if global and local hydroclimatic processes can be used to predict algal magnitude, severity, and duration in each lake using a statistical modeling and forecast validation (hindcast) approach.

Specifically, we address the following questions:

1) Do local and global (sea surface temperature, SST) hydroclimatic variables correlate with chlorophyll metrics in a given lake?
2) Are skillful predictions based on these variables possible for algal magnitude, duration, and severity?
3) Can variability in forecast model performance be explained by static, lake-specific characteristics?

This modeling approach may provide insight into the role of local and global hydroclimatic variability in the development of peak season algal biomass, evaluate the ability of hydroclimatic variables to provide actionable information to lake managers at a seasonal timescale, and indicate which characteristics of small inland lakes make them ideal candidates for seasonal forecast development.

2. Materials and methods

2.1. Lake characterization and selection

Chlorophyll-*a* measurements, sampled at the surface of the lake and analyzed following project-specific protocols for each lake, were obtained from LAGOS-NE (Soranno et al., 2017). The measurements range from 1982 to 2013 in the database, but several chlorophyll timeseries were extended through 2020 by collating additional measurements from the reporting agency or program referenced in LAGOS-NE for each lake. LAGOS-NE aggregates data from lakes located throughout the Midwest and Northeast U.S. In this region, chlorophyll-*a* tends to reach peak concentrations between July and October (JASO) (Fig. 1). The data from this period were used to calculate three chlorophyll metrics: *magnitude, severity*, and *duration* (see explanation below) for each lake year. To adequately characterize July-October chlorophyll-*a* metrics, sufficiently long observational records and frequent within-season sampling are needed. Therefore, lakes for this analysis needed at least 15 years of July-October chlorophyll-*a* measurements and a minimum sampling frequency of once every 14 days, following the selection methods of Wilkinson et al. (2021). Based on these requirements, 178 lakes were identified from 10 states in the Midwest (Michigan, Wisconsin, Minnesota, Ohio, and Missouri; 64 lakes) and Northeast (New York, Vermont, Rhode Island, Pennsylvania, and Maine; 114 lakes) U.S. The chlorophyll data were log-transformed for analysis to create a Gaussian-like distribution. Selected lakes had an average depth of 15.8 m (min=1.2 m, max=198.4 m) and an average area of 1297.9 hectares (min=1.4 ha, max=113,496.5 ha).

2.2. Chlorophyll metrics

To characterize (and eventually predict) algal conditions in each lake year during the July-October season, three metrics were used: *magnitude, severity*, and *duration* of chlorophyll-*a,* as defined by Wilkinson et al. (2021). *Magnitude* is the mean chlorophyll-*a* concentration in each lake year, and *duration* is the portion of the season during which chlorophyll-*a* concentrations exceed a threshold concentration for recreational value based on Angradi et al. (2018). Here, the *severity* metric has been altered from Wilkinson et al. (2021) and is defined as a function of *magnitude*. Seasons in which *magnitude* exceeds the 95th percentile of all historical chlorophyll concentrations (rather than year specific concentrations) are categorized as *severe.* On an inter-annual timescale, *magnitude* characterizes the average conditions during the peak season and the corresponding impacts on ecosystem services. *Severity* reflects the probability of extreme algae biomass (*magnitude)* that is most likely to result in severe consequences like toxin production and fish kills. Finally, *duration* is the persistence of high algal biomass associated with a loss of recreational value during the summer, peak season. The threshold concentrations used here are developed for two ecoregions (“Mountains” and “Plains”) and vary in each region based on recreational user’s expectations for water quality. Because these are large regions, lakes that fall below or above the chlorophyll-a threshold in nearly all sampling events are removed from analysis to avoid artificially inflating overall forecast skill. Together, these metrics may provide an enhanced understanding of algae conditions during the peak season of algal production, with prospects for more refined actionable information. Compared to a singular forecast of mean chlorophyll (magnitude), a forecast that additionally provides advanced warning of protracted water quality impairment (duration) and the potential for severe consequences of algae growth (severity) allows for more nuanced decision making around budgeting, testing, and communicating water quality expectations with the public. The extent to which these three metrics are correlated with local and global climate variables and predictable across a diverse set of lakes is the focus of this work.

Table 1
Variables used in correlation analysis with peak season chlorophyll-*a* metrics, including data source and resolution.

Predictors (March-June)	Source	Resolution
Total precipitation (mm)	PRISM	4km
Mean air temperature (°C)	PRISM	4km
Mean volumetric soil moisture (m^{-3}/m^{-3})	Copernicus Climate Change Service	0.25°
Precipitation events exceeding 20 (40) mm in Midwest (Northeast) watersheds	PRISM	4km
Sea surface temperature anomalies (°C)	NOAA ERSST	2°
Pre-season Chlorophyll-*a* (µg/L)	LAGOS-NE	In-situ

2.3. Predictor variable selection

To address whether local and global hydroclimatic variables are correlated with chlorophyll metrics in a given lake, we evaluate to what extent pre-season (March through June) observations of local and global hydroclimate variables are correlated with magnitude and duration for each lake. Correlations for *severity* were not evaluated independently in this analysis as the *severity* metric is a function of *magnitude*. In addition to identifying common local and global hydroclimatic variables correlated with chlorophyll metrics, this analysis was used to identify variables that would be used in the development of each lake-specific forecasting model, and the validation of each model in a hindcasting analysis. The hindcasting analysis uses the lake-specific forecasting model to predict each year in the chlorophyll metric timeseries without predictor information from the year of interest, simulating a forecast for model validation. Specifically, we included March-June variables from readily available, gridded datasets that were connected to physical processes that may affect *magnitude* and *duration* including (Table 1): total precipitation (mm), mean air temperature (°C), mean volumetric soil moisture (m^{-3}/m^{-3}), the sum of daily precipitation events exceeding 20 (40) mm for Midwest (Northeast) watersheds, and global

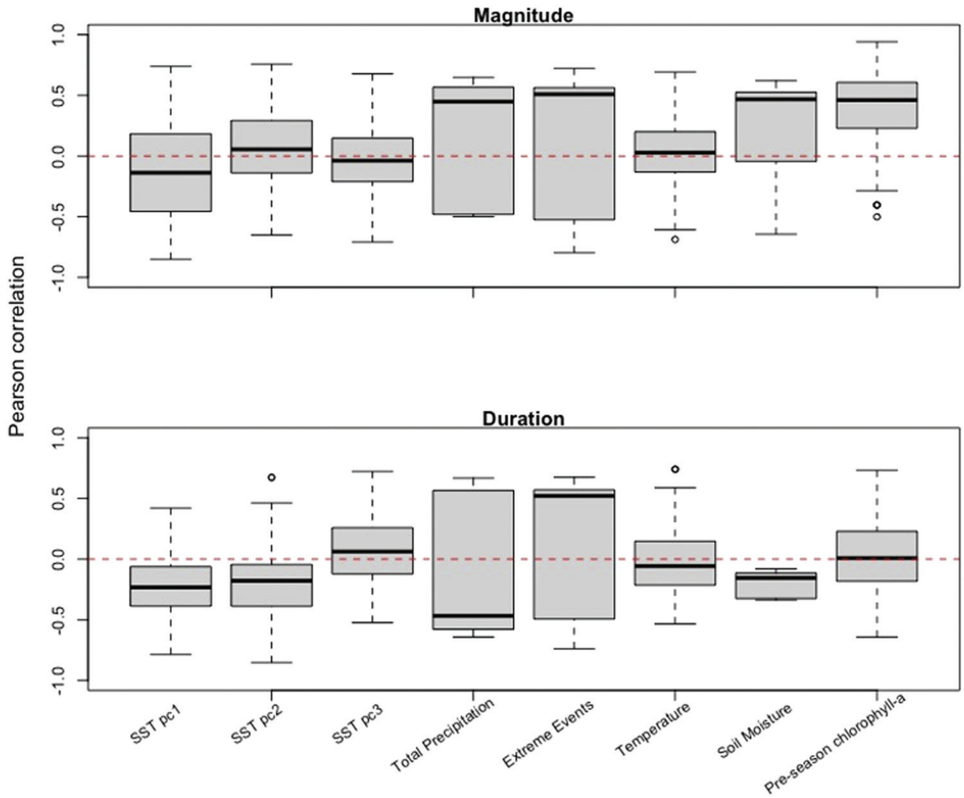

Fig. 2. The distribution of Pearson correlations between candidate predictors and algae metrics for all lakes.

sea surface temperature (SST) anomalies (°C). Methods for evaluating SST anomalies as predictors are described below. Additionally, we evaluated if pre-season chlorophyll-a, which reflects in-situ processes and the nutrient availability at the start of the season, is correlated with peak-season chlorophyll metrics. Excluding extreme events, all metrics were averaged over the March-June season.

As discussed previously, local hydrology regulates nutrient transport into lakes from the surrounding watershed, which may ultimately influence algae abundance. The influence of local hydrology-based candidate predictors (precipitation, extreme events, and soil moisture) may vary based on land use and topography within a lake's watershed. Therefore, local hydrology predictors were evaluated using a correlation analysis at each grid within each lake's HUC12 watershed. HUC12 watershed polygons (Watershed Boundary Dataset, 2021) were subset to only include areas higher in elevation than the corresponding lake (Fig. A.1) using gridded elevation data for each watershed from the elevatr package for the R statistical programming language (Hollister et al., 2021). The timeseries of candidate predictor variables from each grid intersecting the watershed polygon was used in the correlation analysis as was an average of all intersecting grids. The grid with the strongest, statistically significant correlation was retained for the subsequent hindcasting analysis. High precipitation events may have a more significant influence on overall nutrient loading than total precipitation (Carpenter et al., 2015), however, precipitation events that lead to large loading events may vary by region due to land use topography, and nutrient availability. Therefore, separate thresholds were chosen for extreme precipitation events in Midwest (20 mm) and Northeast (40 mm) watersheds based on a sensitivity analysis of significant correlations between high precipitation events and peak-season chlorophyll-*a magnitude* conducted for each region.

On a global scale, pre-season sea surface temperatures can influence in-season precipitation and temperature over the U.S. through modulation of atmospheric flow and thus indirectly influence peak-season chlorophyll metrics (Barnston, 1994; Giannini et al., 2000; Markowski and North, 2003). SSTs evolve slowly, with persistent (months to years) anomalies, and thus can serve well as predictors at seasonal timescales (Barnett, 1981). To identify oceanic regions with strong teleconnections to the Northeast and Midwest U.S., global pre-season sea surface temperature (SST) anomaly grids were correlated with each lake's magnitude timeseries (Fig. A.2). Not surprisingly, correlation patterns, and thus oceanic regions of influence, vary between the Midwest and Northeast U.S. (Ropelewski and Halpert, 1987; CPC, 1997; Enfield et al., 2001; Tootle et al., 2005), therefore identification of teleconnections is performed separately for the Midwest and Northeast. The number of significant correlations with chlorophyll-a timeseries were tallied for each SST grid and mapped to identify oceanic regions in which SST grids were associated with algae abundance in the Northeast and Midwest. SST grids that were significantly correlated with a large fraction of lakes (≥10 for Midwest, ≥20 for Northeast) were applied to a principal component analysis (PCA) to extract the dominant modes of variability in SST data and reduce the dimensionality of candidate SST predictors. Given the large number of SST grids, there is a high likelihood of generating spurious correlations. Performing PCA extracts the dominant climate signals and minimizes the effect of spuriously correlated grids. Principal components (PCs) that explained more than 5% of the variance in SST anomaly data were retained as candidate predictors.

The time series of variables in Table 1 were used in a correlation analysis with chlorophyll *magnitude* and *duration* for each lake to identify predictor variables for forecast model development and the subsequent hindcasting analysis (Fig. 2). For each lake, all variables that were significantly correlated with the chlorophyll metric ($P<0.05$) were retained for the forecasting model. Out of the 178 lakes evaluated, 135 lakes (50 Midwest; 85 Northeast) had at least one significant predictor variable for magnitude and 82 lakes (30 Midwest; 52 Northeast) for duration. *Severity* is a function of *magnitude* and therefore retained the same set of predictors.

2.4. Forecast model development

Two forecasting models were developed for each lake, one focused on *magnitude (including severity*) and a separate model for *duration* (proportion of sampling events above the impairment threshold). The array of processes and feedbacks influencing algae growth and abundance are notoriously complex (Roelke and Buyukates, 2001; Glibert and Burkholder, 2006; Ho et al., 2019), motivating a statistical modeling approach over a process based/physical model approach. For lakes with only one significant predictor from the correlation analysis, a simple linear regression between that variable and the chlorophyll metric was constructed for the model to be used in the hindcast analysis. For lakes with multiple significant predictors, a principal component analysis and regression approach was used to build the forecasting model. PCA effectively deals with any multi-collinearity present between predictors and therefore does not artificially inflate predictive skill. Here, principal components were retained for the forecast model if they explained more than 10% of the variance. This modeling approach assumes relationships between candidate predictors and peak season algae metrics on a seasonal timescale to be linear, however, given that many drivers of algae growth on short timescales (days to weeks) are considered nonlinear processes, model residuals were evaluated for evidence of nonlinear relationships. Autocorrelation was also investigated in each of the candidate predictors and algae metrics. Except for SST PC1, which likely captures baseline increases in the temperature of the Pacific Ocean, less than 10% of timeseries for each variable had more than two statistically significant autocorrelations (lag 1–10). Additionally, random forest regression, a nonlinear, nonparametric modeling approach, was tested to determine if there were notable changes in model skill due to potential nonlinearities or autocorrelation. Statistical models were developed using R version 4.2.1.

A leave one out cross-validation approach was used to evaluate model performance for each lake and chlorophyll metric (hindcasting). In short, for each lake the observed value from one year of the chlorophyll metrics was removed from the timeseries and the forecasting models (above) were used to predict the missing value. This process was done iteratively for all years in the timeseries for all lakes individually and both *magnitude* (including *severity*) and *duration*. A prediction ensemble was created for each peak-season chlorophyll metric in a lake year based on model errors (difference between observed and predicted chlorophyll metric) across the hindcast at that lake. Ensemble members are generated from a normal distribution of errors with mean zero, based on maximum likelihood estimation. For each time-step, 100 random draws from the distribution are added to the *magnitude* and *duration* predictions to form the ensemble prediction (Helsel and Hirsch, 1992; Alexander et al., 2019).

2.5. Performance measures

To assess model performance, four measures were adopted: correlation coefficient (R^2), root mean square error (RMSE), ranked probability skill score (RPSS), and Heidke skill score (HSS) (Heidke, 1926; Epstein, 1969). HSS and RPSS are measures of categorical skill, interpreted as a percent improvement over a reference forecast. A standard forecast for hydro-climate prediction is an equal-odds (climatological) distribution of historical observations. Here, the distribution of historical observations of chlorophyll magnitude for each lake are split into four categories representing *below normal, near normal, above normal,* and *severe* algae conditions. If no predictive information is present, a probabilistic prediction of JASO magnitude would default to climatology (33% chance of *below normal*, 33% chance of *near normal* conditions, 28% chance of *above normal*, and 5% chance of *severe*

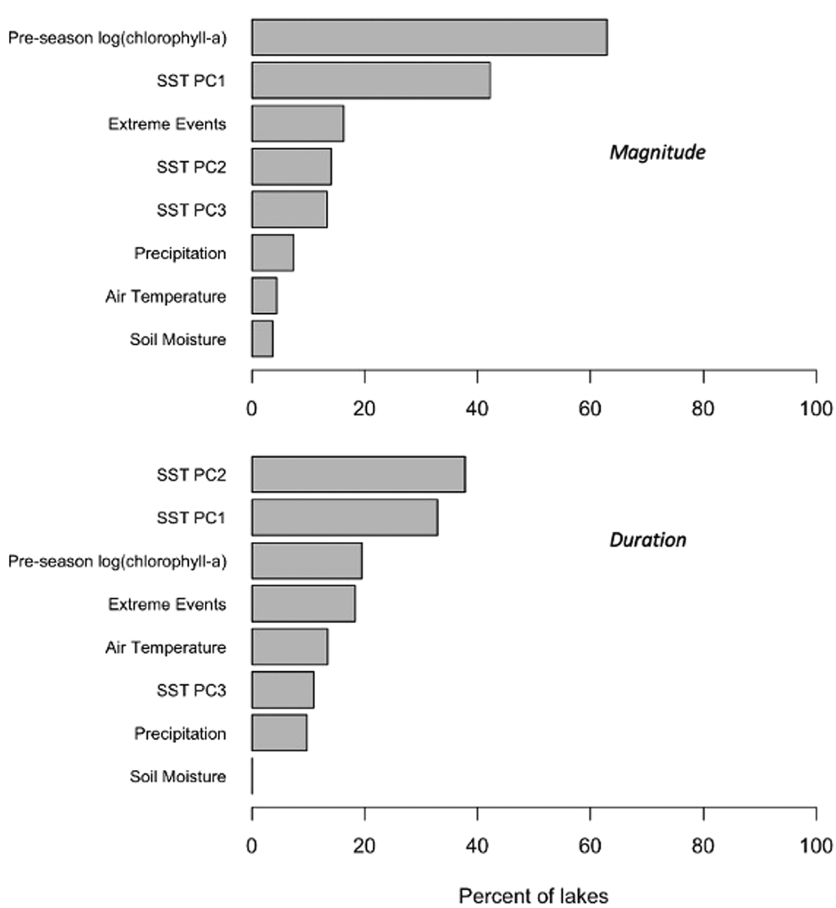

Fig. 3. The percent of lakes for which each predictor is statistically significantly correlated with *magnitude* (top) and *duration* (bottom); 178 lakes evaluated.

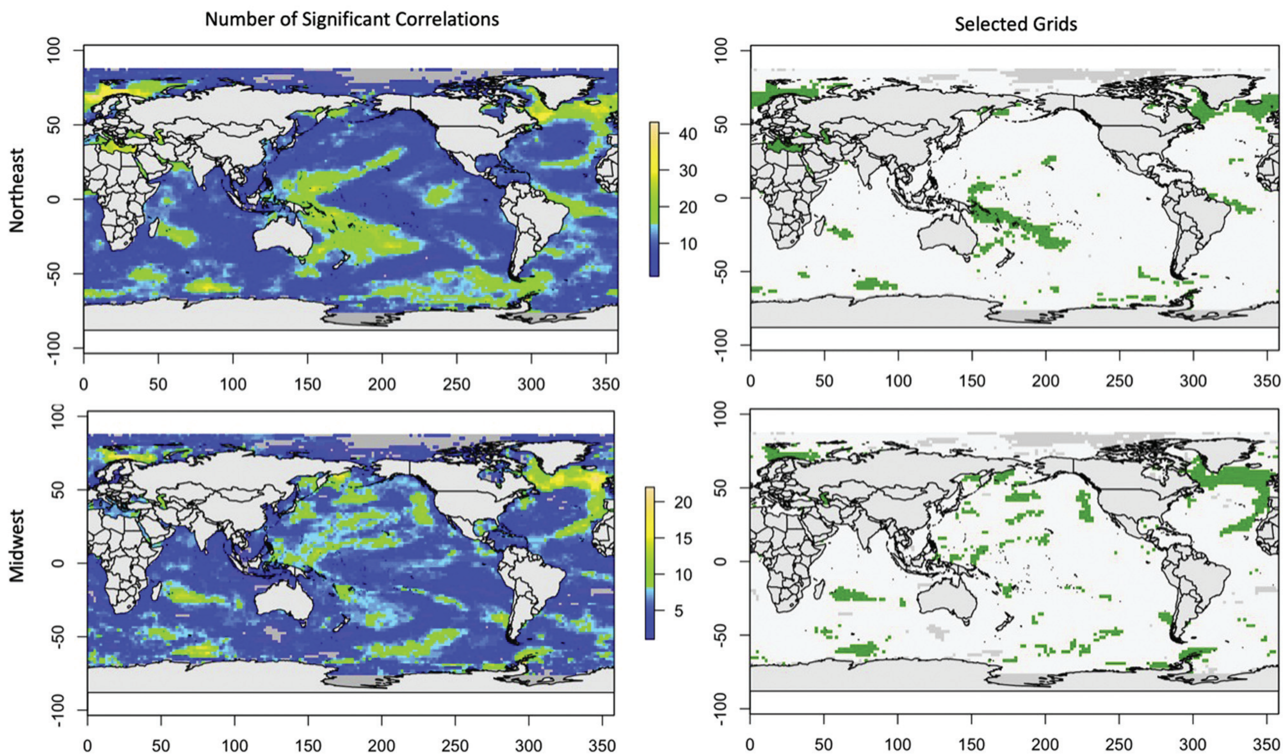

Fig. 4. Number of preseason SST grids that are statistically significantly correlated with chlorophyll-*a* timeseries (left column) from lakes in the Northeast (top row) and Midwest (bottom row). Grids retained (right column) have at least 20 (10) significantly correlated lakes for the Northeast (Midwest).

conditions). Similarly, observations of duration are split into two categories based on mean duration to represent expected below and above normal conditions. Forecast models developed for magnitude and duration generate probabilistic predictions of each category that are compared against climatology (equal odds). This allows for a direct comparison between the forecast models developed here and a benchmark climatology model to understand the prospects for enhanced predictive skill. In general, prediction models outperform climatology when the predicted probability of the observed category is greater than the climatological probability (e.g. 50% for two categories, 33% for three categories). HSS is defined as:

$$HSS = \frac{H - E}{N - E} \tag{1}$$

Where H is the number of categorically correct forecasts, N is the total number of predictions issued, and E is the number of categorically correct predictions expected from the reference forecast. HSS values range from -∞ to 1, where negative values represent a forecast that performs worse than climatology, 0 represents no skill, and 1 represents a perfect prediction model. The RPSS is a categorical skill score that increasingly penalizes an ensemble forecast for assigning greater probability to categories farther from the observed category. The RPSS uses the ranked probability score (RPS), the average of the squared difference between the cumulative probability of the forecast and observations (Eq. (2))

$$RPS = \sum_{i=1}^{n} (CPfct_i - CPobs_i)^2 \tag{2}$$

where $CPfct_i$ and $CPobs_i$ are the cumulative probabilities of the forecast and observed values through categoIy *i,* and n is the total number of categories. The RPSS is then defined as:

$$RPSS = 1 - \frac{RPS_{forecast}}{RPS_{climatology}} \tag{3}$$

Where $RPS_{forecast}$ and $RPS_{climatology}$ are the RPS values calculated using the forecast model and the reference forecast. RPSS values range from -∞ to 1, where 0 represents no skill and 1 represents a perfect forecast. RPSS values are calculated for each year and the median value is reported.

Finally, we evaluated if variability in hindcast model performance (forecasting skill) among lakes was related to static characteristics of the ecosystems. We compared forecasting skill among categories of trophic state, lake area, land cover, and geographic region among lakes. The trophic state index (TSI) is calculated based on chlorophyll-*a* and is categorized as oligotrophic (TSI<40), mesotrophic (40≤TSI<50), eutrophic (50≤TSI<70), and hypereutrophic (TSI>70) (Eq. (4)) (Carlson, 1977).

$$\mathrm{TSI(CHL)} = 9.81\ \ln(\mathrm{CHL}) + 30.6 \tag{4}$$

3. Results

3.1. Leading algae characteristic predictors

The three most frequently retained predictors include pre-season chlorophyll-*a*, PC1 from SSTs, and extreme events for *magnitude* models. The most frequently retained predictors for *duration* models include pre-season chlorophyll-*a*, and PC1 and PC2 based on SSTs. PCs derived from SSTs typically represent dominant large-scale climate signals, potential physical processes are explored further in the discussion. *Magnitude* and *duration* metrics are uncorrelated in most lakes (only 11are significantly correlated at the 95% confidence level), suggesting unique seasonal drivers for each metric in most lakes. Compared to *magnitude* models, *duration* models have a more even distribution of retained predictors (Fig. 3). In *magnitude* models pre-season chlorophyll-*a* meets the selection criteria in 63% of models, SST PC1 in 42% of models, and extreme events in 16%. In *duration* models, SST PC2 is selected in 38% of models, SST PC1 in 33%, and pre-season chlorophyll-*a* in 20%. In both *magnitude* and *duration* models, all three predictors are

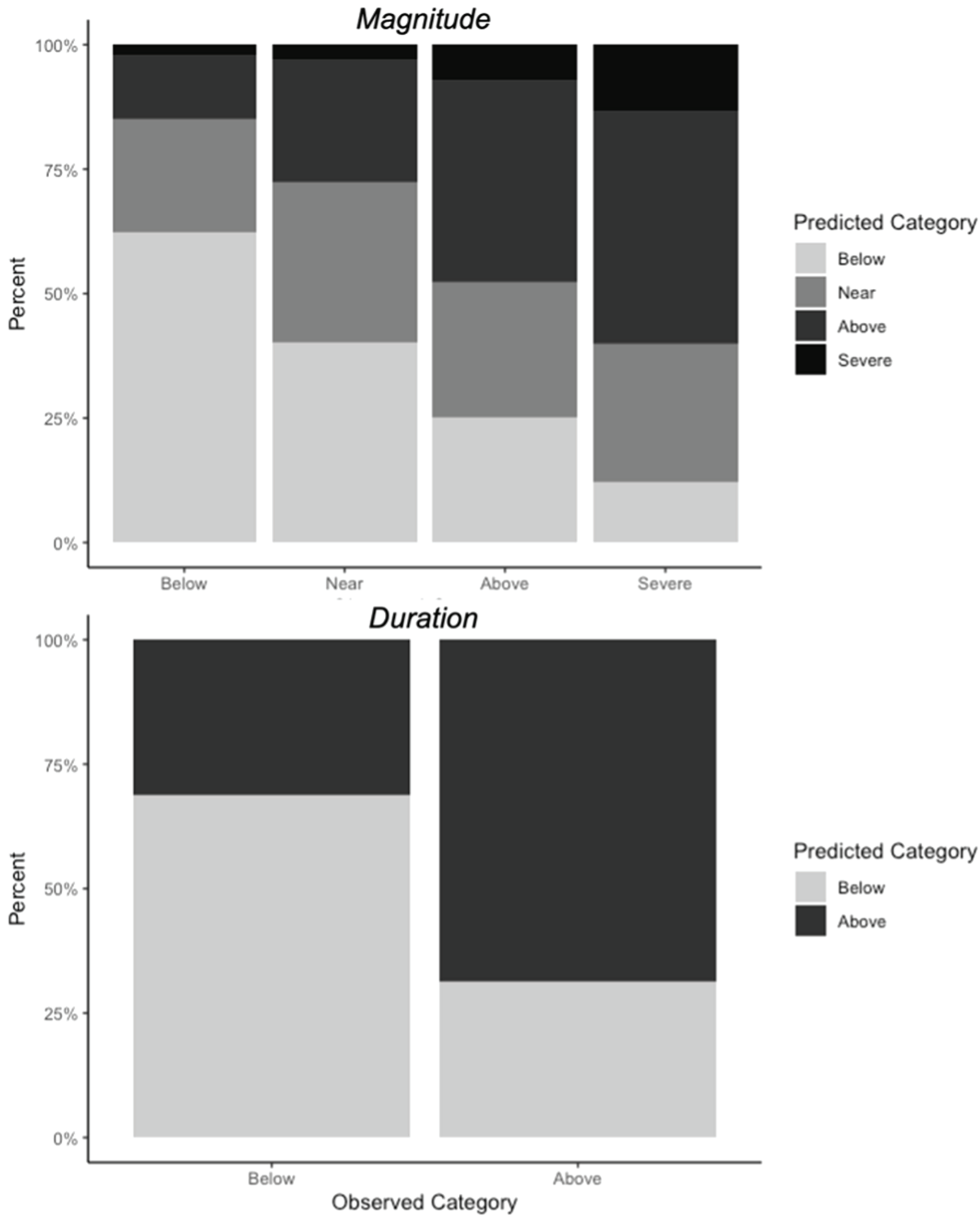

Fig. 5. The proportion of each prediction category for each observed category (all lakes).

unlikely to appear in the same model, suggesting variable influence of these processes by lake. In *magnitude* models, all three of the most frequently retained predictors are included in 7% of lakes. In contrast, one of the three predictors in included in 87% of *magnitude* models. In *duration* models, all three of the most frequently retained predictors are included for only two lakes (2.5%), while one of the three predictors is included in 71% of *duration* models.

Preseason SST grids that are statistically significantly correlated with chlorophyll-*a* timeseries from at least 20 (10) lakes are retained for the Northeast (Midwest) region (Fig. 4). SST regions retained for both Midwest and Northeast lakes indicate teleconnection signals from the northern Atlantic and the equatorial pacific. For Midwestern lakes, more grids are retained in the mid and northern Atlantic compared to the Pacific. Comparatively, the strongest signals for northeastern lakes are split more evenly between the upper Atlantic and the equatorial pacific.

While local hydroclimatic predictors are present in a significant proportion of *magnitude* and *duration* models, in most instances they are retained with SSTs or pre-season chlorophyll-*a,* particularly in *magnitude* models. Some combination of local hydroclimatic predictors (precipitation, extreme events, soil moisture, and air temperature) are retained in 27% of *magnitude* prediction models, while just 9% *of magnitude* models utilize only local hydroclimatic predictors (i.e., without SST PCs or preseason chlorophyll-*a).* Compared to *magnitude* models, a larger proportion of *duration* models are built solely with local hydroclimatic variables. Hydroclimatic predictors are retained in 38% *duration* models, and 22% of models are built exclusively with hydroclimatic predictors

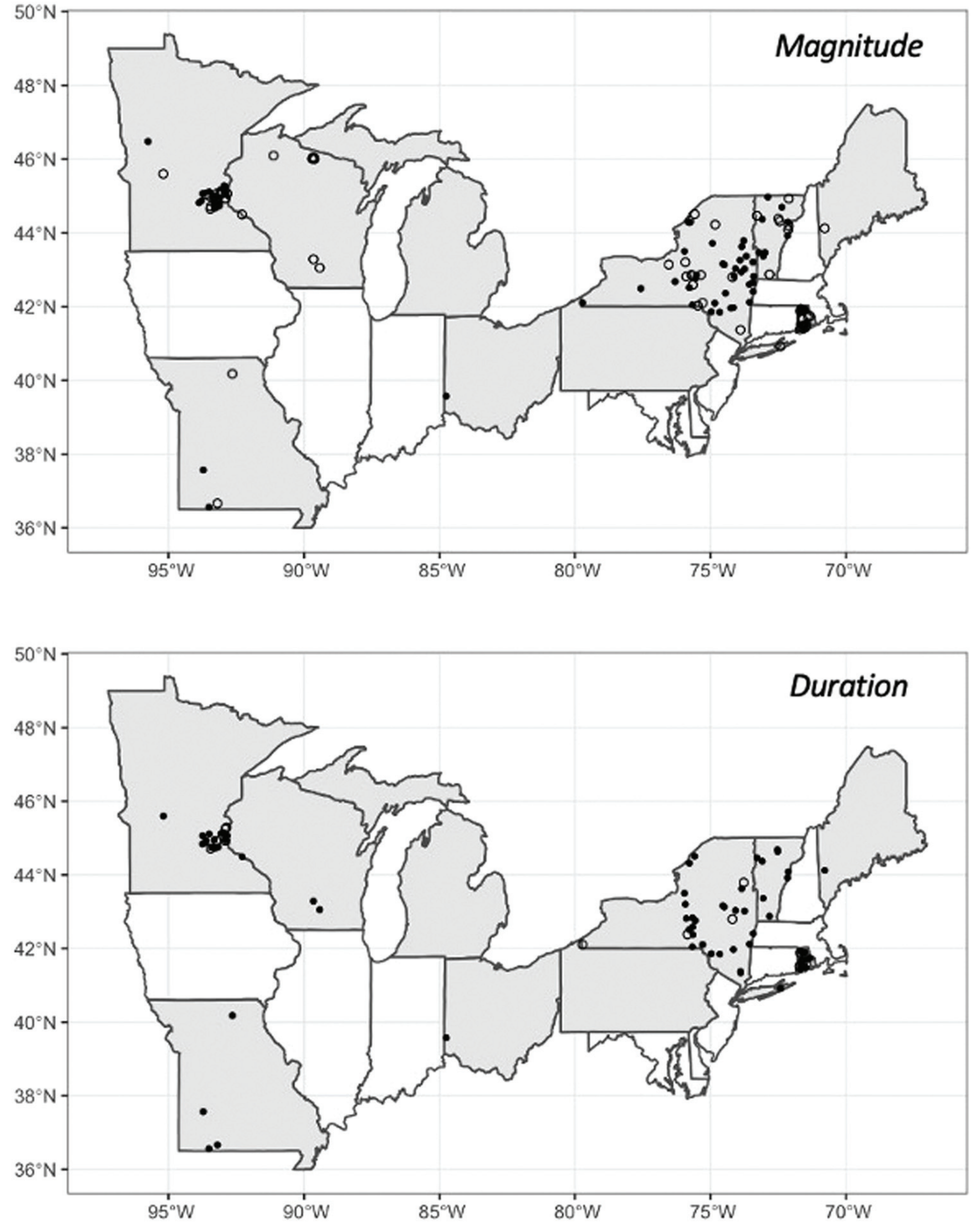

Fig. 6. Lake models that correctly predict more (less) than half of elevated *magnitude* (top) and *duration* (bottom) events, illustrated as open (closed) circles.

(Fig. A.3). Finally, 86% of lakes have 1 or 2 significant predictors while the maximum number of predictors retained for a lake is 5 (Fig. A.4).

3.2. Model performance

For *magnitude* prediction models, simple linear regression (single predictor) is applied to 53% of lakes, whereas principal component regression (PCR, principal component analysis with multiple linear repression) is applied to 47% of lakes. A cross-validated hindcast assessment for all lakes results in a mean R^2 value of 0.28 (0–0.85) and a mean RMSE of 0.47 (0.29–1.21). For categorical performance, the mean HSS and RPSS values are 0.17 and 0.10, respectively. Additionally, 87% (70%) of lake models have HSS (RPSS) values greater than zero, indicating an improvement of prediction skill over climatology for most lakes. Further, these models predict *above normal* and *below normal* algae abundance moderately well (Fig. 5), with RPSS values of 0.39 and 0.30 respectively. *Severe* events prove difficult to predict; approximately 51% of *magnitude* models accurately predict an *above normal* or *severe* year when an elevated algae event is observed (i.e., not *below* or *near normal*) for more than half of observed elevated algae events (Fig. 6).

For *duration* prediction models, simple linear regression is applied to 78% of lakes and PCR to 22% of lakes. Mean R^2 and RMSE are 0.23 (0–0.65) and 0.34 (0.16–0.42), respectively; mean HSS and RPSS scores are 0.39 and 0.42, respectively, and 96% (94%) of models improve over climatology based on HSS (RPSS). In a hindcast assessment, *durations* of *above normal* are correctly predicted for more than half the available timeseries in 87% of lakes (Fig. 6). Additionally, considering all models, 67% of *above normal durations* are accurately predicted (Fig. 5), a stark improvement over climatology for most lakes.

Average model skill is similar between regions is similar, however, Northeast lake forecast models outperform Midwest lake models where differences in average skill scores occur (Fig. 7). This may be expected given that the primary difference in predictor selection between models in the Northeast and Midwest is selection of relevant SST grids. Relevant Northeast SST regions are more coherent than regions for the Midwest, which may represent a stronger influence on lake processes. The minimal differences may also point to consistency in the predictive power of local and within-lake variables between the Northeast and Midwest.

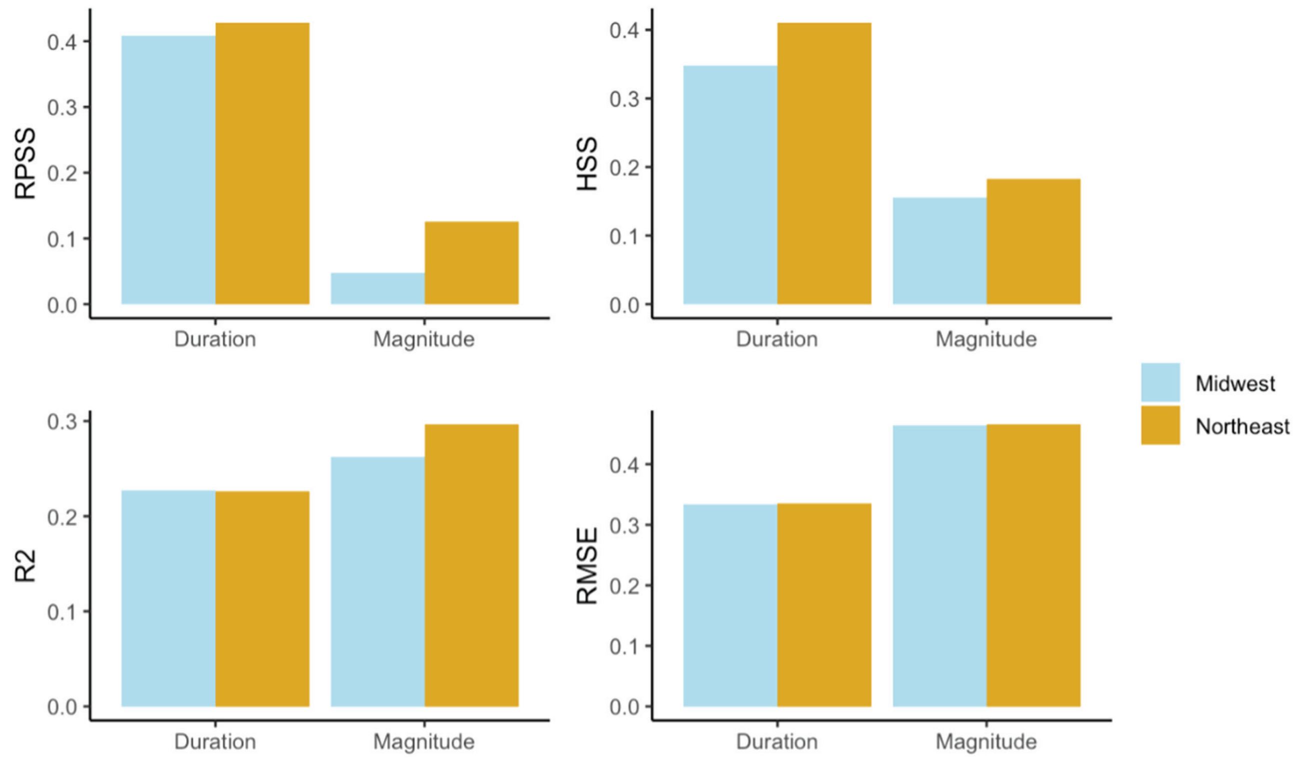

Fig. 7. Average model skill scores for *duration* and *magnitude* models by region.

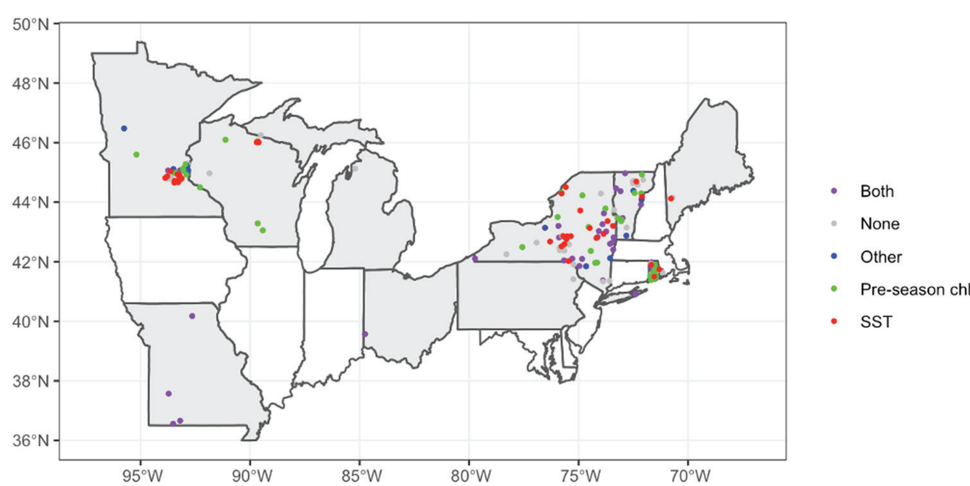

Fig. 8. Lakes in which chlorophyll-*a*, SSTs, both, or neither are selected for *magnitude* predictions.

To evaluate the presence of nonlinear relationships, residuals between predicted and observed chlorophyll metrics were investigated for each lake. Residuals generally appeared random, suggesting that the relationships between the pre-season drivers and peak season chlorophyll metrics investigated for this analysis can be approximated as linear. Hindcasts were also generated using random forest regression to test for an increase in predictive skill, which may indicate the presence of nonlinear relationships. Random forests models were created with the same predictors selected for PCR, each with 500 trees. Cross validated hindcast results are similar or slightly worse than PCR for both *Magnitude* (Mean: $R^2 = 0.25$, RMSE = 0.48, HSS = 0.12, RPSS = 0.04) and *Duration* (Mean: $R^2 = 0.23$, RMSE = 0.34, HSS = 0.34, RPSS = 0.39), suggesting that PCR is a suitable approach.

4. Discussion

4.1. Regional characteristics

Predictions of peak season algae growth at scale provide insights into relevant global and local-scale processes setting the conditions for peak season algae biomass. Pre-season SSTs and chlorophyll-*a* observations provide the most predictive power for both *magnitude* and *duration* metrics, reflecting the importance of these scales. However, while SSTs and chlorophyll-*a* are selected most frequently as predictors for both *magnitude* and *duration*, the importance of each predictor is mixed by region (Fig. 8). Given that SST-atmosphere teleconnections typically have regional influence, spatial variability in performance of SST predictors may be the result of localized pre-season processes superseding the influence of large-scale climatic processes in peak season algae

Table 2
Average skill scores for *magnitude* and *duration* models in which pre-season chlorophyll-*a* and/or SSTs are utilized as predictors.

Variable	High/Severe Correct	RPSS	HSS	R^2	RMSE	Predictand
Both	0.58	0.22	0.20	0.43	0.49	*Magnitude*
MAMJ chlorophyll-*a*	0.47	0.03	0.15	0.24	0.50	
SSTs	0.51	0.08	0.16	0.24	0.40	
Other	0.44	0.08	0.14	0.17	0.43	
Both	0.65	0.55	0.40	0.31	0.33	*Duration*
MAMJ chlorophyll-*a*	0.56	0.12	0.10	0.18	0.36	
SSTs	0.70	0.43	0.41	0.21	0.34	
Other	0.67	0.22	0.27	0.11	0.33	

biomass. For example, food web dynamics are well established as having significant influence on aquatic primary productivity (Lampert et al., 1986; Carpenter et al., 1987; Vanni and Temte, 1990). In this study, however, the representation of food web dynamics is limited to pre-season algae abundance. While this variable is shown to be a powerful predictor of peak season productivity in many lakes, the effect of pre-season predatory control on algae communities is unrepresented. This may limit the skill of prediction models for lakes in which zooplankton grazing plays a significant role in determining algae populations in the summer and early fall. This limitation may be responsible for differences in average model performance. For example, *magnitude* and *duration* models in which pre-season chlorophyll-*a* is an important predictor, and SSTs are not, perform worse on average than the inverse (Table 2). Lakes retaining only pre-season chlorophyll-*a* may be more dependent on within-lake processes, many of which are not represented in these models. This may explain the less robust predictive signal from pre-season chlorophyll-*a* in these lakes, compared to lakes more heavily influenced by SST-atmosphere teleconnections.

As discussed previously, relevant SST anomaly grids are identified for the Northeast and Midwest separately. The PCs of selected SST anomaly grids represent the dominant climate signals affecting the selected lakes across the Northeast and Midwest U.S. and may therefore be associated with large-scale climate phenomena that have well-established teleconnections with climate conditions across North America. Two dominant climate phenomena with variable impacts on local hydroclimatic conditions across the Northeast and Midwest include the North Atlantic Oscillation (NAO) and the El Niño Southern Oscillation (ENSO) (Ropelewski and Halpert, 1987; Visbeck et al., 2001). In the Northeast, two of the three SST PCs used as potential predictors are significantly correlated with both the NAO index and multivariate ENSO index (MEI). In the Midwest, two PCs are significantly correlated with the MEI and all three are significantly correlated with the NAO index. This suggests that interannual variability in local climate, and in lakes, across the Northeast and Midwest is associated with both ENSO and NAO-like climate signals (Fig. A.5).

Compared to pre-season chlorophyll and SSTs, other variables considered here provide only modest skill in predicting algae characteristics for most lakes. Despite the perceived importance of local land and hydrologic variables modulating inflow and lake processes, few were retained as predictors; however, land use and other watershed characteristics may be important in determining the relevance of these predictors. Local hydrology might be expected to play a larger role in promoting algae growth in agricultural watersheds, given the effect of runoff on nutrient loading (Castillo et al., 2000; Mander et al., 2000). This is reflected in predictor selection for model construction, for example, *magnitude* models in watersheds with greater than 25% agricultural land are nearly twice as likely to retain a local hydrologic predictor compared to models in watersheds with less agricultural land (18.4% vs. 36%). March-June air temperatures are also retained in

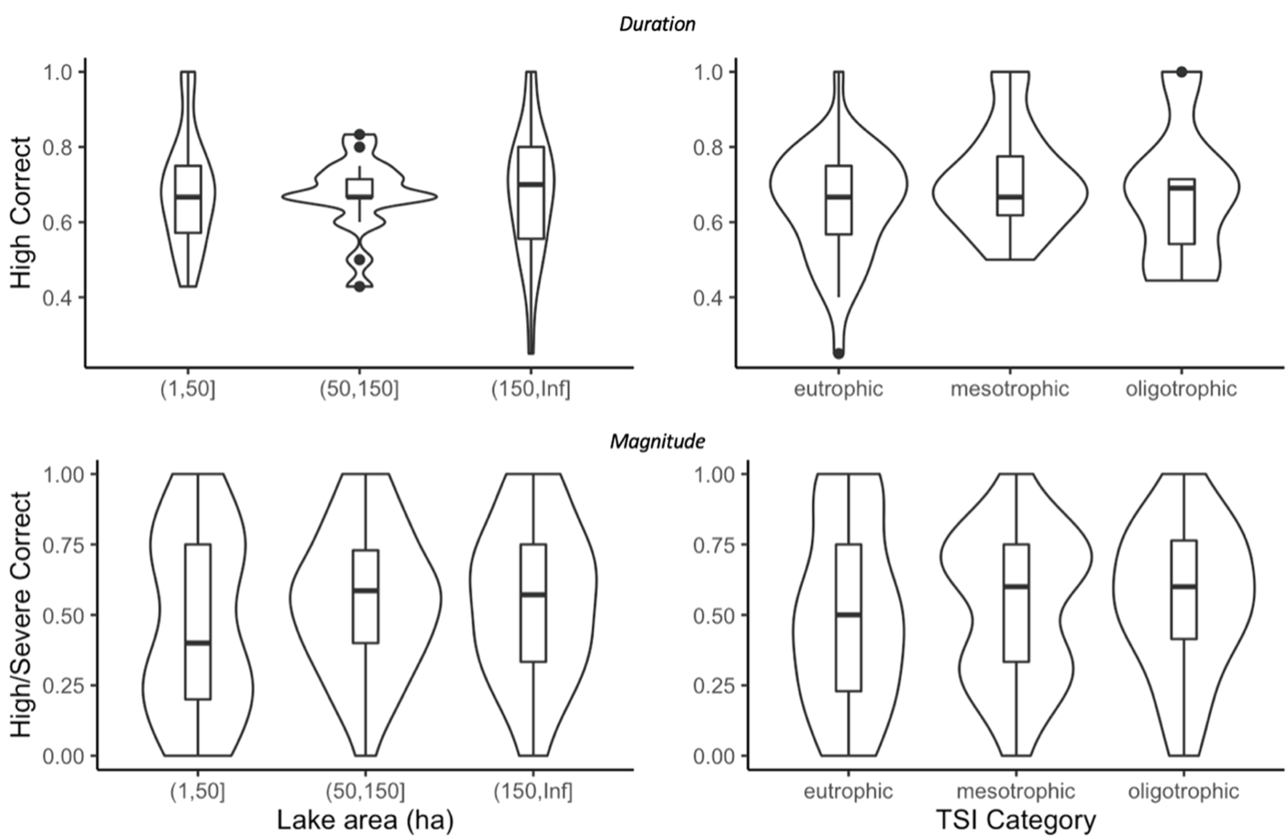

Fig. 9. Violin plots of the proportion of *above normal* or *severe* events correctly predicted for each algae metric, compared to lake area (ha) and Trophic State Index (TSI).

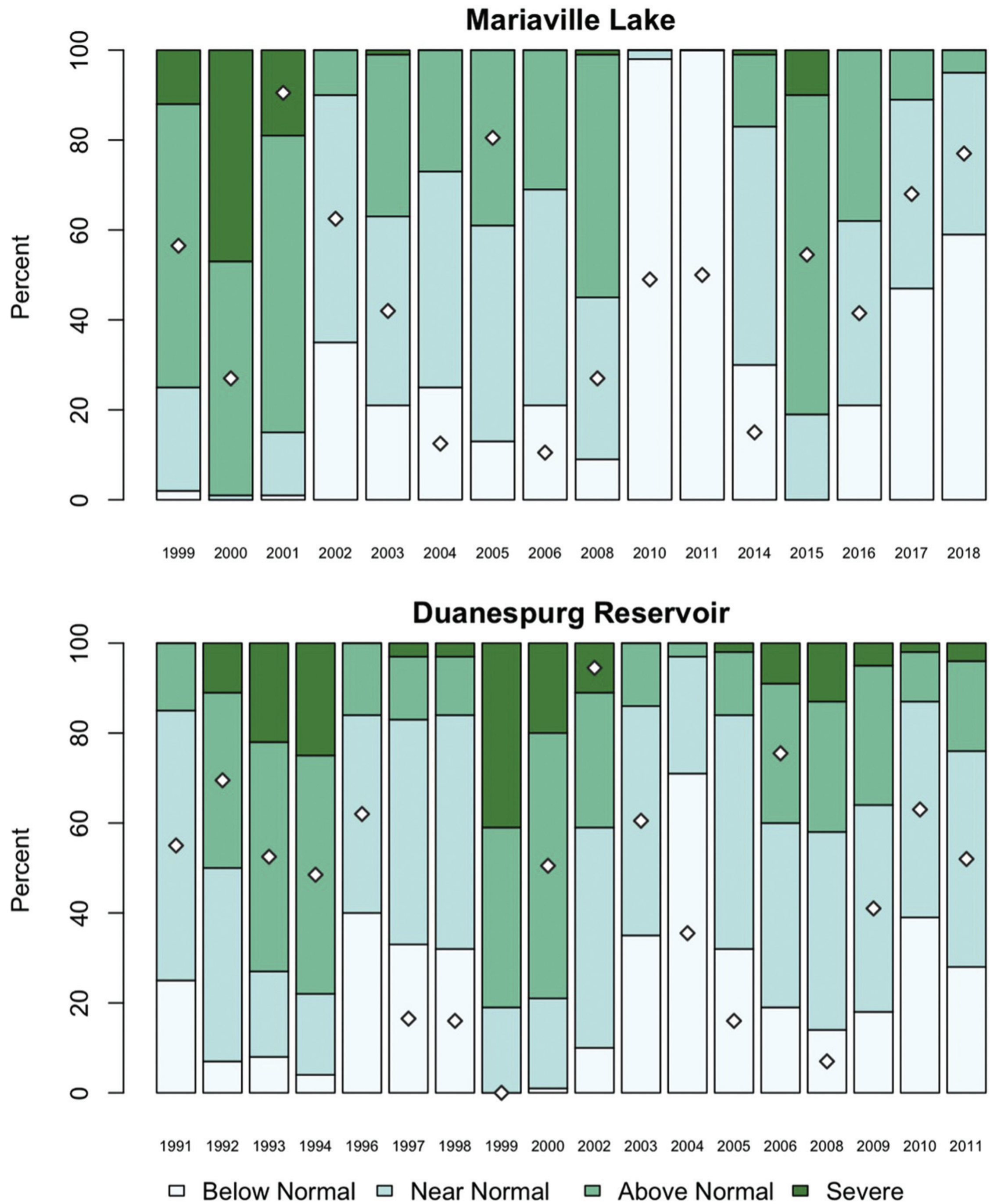

Fig. 10. Probabilities of *magnitude* prediction categories for Mariaville Lake and Duanespurg reservoir, New York. Diamonds indicate the observed category.

relatively few forecasting models overall but are notably included in more *duration* models than any of the local hydrologic predictors. As discussed previously, the importance of temperature in determining algae growth is well established, however, the air temperature predictor included here may be too simplistic to capture the relationship across all lakes. Significant variability in prediction skill exists among lakes with little evidence of spatial patterns. On average, skill scores are higher in the Northeast compared to the Midwest, however, variability within both regions is significant. The frequency of predictor retention and the average magnitude of significant correlations between predictor variables and algae *magnitude* are similar by region particularly for the most frequently selected predictors, including SSTs, pre-season chlorophyll-a, and extreme events (within 0.05). SST PC1 and extreme events in the Northeast have slightly higher correlations with algae *magnitude* than in the Midwest, which may help explain slightly higher model skill scores in the Northeast. The magnitude of correlation between predictor variables and algae *duration* is also similar, however, SST PC2 is retained much more often in Northeast models compared to Midwest models (25% of lakes vs 5% of lakes). Given that SST PC2 is correlated with the NAO, this might indicate a greater influence of the NAO on *duration* in the Northeast and help explain the slightly higher skill scores of *duration* models in the Northeast.

Similarly, distributions of skill scores across trophic state and lake area are variable (Fig. 9).

On average, magnitude models appear to accurately predict increased algae abundances more frequently in larger lakes and in

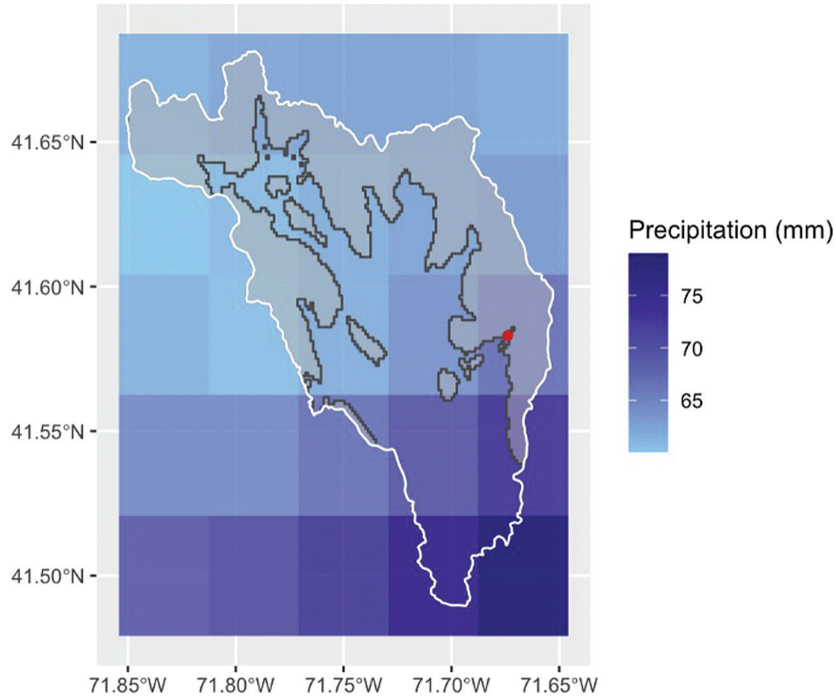

Fig. A.1. An example of the grid selection process for hydrologic predictors. The white polygon represents the HUC12 watershed, the red dot represents the lake, and the gray represents areas of the watershed with elevation exceeding the lake surface.

mesotrophic and oligotrophic lakes. Duration models have approximately equal distributions of skill across lake area and trophic state. While the differences in magnitude model skill based on lake area and TSI category are notable, they were found to be statistically insignificant in an analysis of variance (ANOVA; lake area $P = 0.24$, TSI $P = 0.36$), therefore it may be difficult to draw definitive conclusions from these results. There are a few potential explanations for the variability in magnitude model performance. In an analysis of lake size and primary productivity in the Canadian Shield lakes, Fee et al. (1994) found that larger lakes more efficiently convert external nutrient loads into phytoplankton biomass due to more frequent resuspension of sediments, and receive a higher proportion of nutrient loads from runoff rather than direct precipitation. The focus in this analysis on hydroclimatic predictors of nutrient runoff is therefore consistent with increased predictive skill in larger lakes. Notably, air temperature did not stand out as an important predictor of algae abundance at a season-ahead lead time. As discussed previously, temperature is well-established as an important variable in algae growth. In some cases warm temperatures have been shown to hold greater influence over phytoplankton growth (Salmaso et al., 2012), and cyanobacteria growth in eutrophic lakes (Rigosi et al., 2014). The air temperature predictor used here may be too simplistic to capture peak season water temperatures, which may disproportionately contribute to lower model performance in eutrophic lakes. Rusak et al. (2018), found a positive relationship between variability of chlorophyll-a and trophic status in 18 globally distributed lakes, which may also reduce predictability (Cottingham et al., 2000). This may explain the moderate reduction in average skill in magnitude models for eutrophic lakes compared to lakes of a lower trophic status.

Variability in prediction skill can also exist between similar or proximal lakes. Mariaville lake and Duansespurg reservoir are two eutrophic waterbodies located in eastern New York, approximately five miles apart. Mariaville lake has one of the best performing *magnitude* prediction models among the lakes considered and accurately predicts *above normal* and *severe* chlorophyll-*a* conditions for each of the five years in which they are observed. Comparatively, Duansespurg reservoir only correctly predicts two out of six observed events (Fig. 10). In 1999, for example, the Mariaville lake model accurately predicts a large probability of *above normal* conditions (observed state) and is even able to differentiate between *above normal* and *severe.* Comparatively, for the same year, the Duansepurg model only predicts a 1% chance of *below normal* conditions (observed state), and both models predict an approximately 80% chance of *above normal* or *severe* conditions in 1999 (77% Mariaville, 85% Duanespurg). The Mariaville lake model includes SSTs (PC1) and pre-season chlorophyll-*a* as predictors, whereas the Duansespurg model includes only SSTs (PC2). The performance of the Duanespurg model compared to the Mariaville model again suggests that while global processes are important in setting conditions for peak season algae biomass, and both explain significant variability in the *magnitude* timeseries of both lakes, within-lake processes that may determine interannual variability of peak season algae abundance are not entirely captured by the hydroclimatic predictors investigated here, or by pre-season algae abundance. The variability in forecast skill and predictive power of hydroclimatic variables among study lakes highlights the importance of catchment- and lake-specific processes and characteristics in determining the effects of external climate forcing on peak-season algae abundance. Catchment soil types, land use, and location, as well as lake area, depth, and management may all influence the susceptibility of lake systems to climate variables (Moss, 2012), and may alter predictive skill.

While forecasting models developed on the selected pre-season predictors cannot entirely capture the nuances of peak season algae biomass, it is notable that relatively simplistic statistical models based on global sea surface temperatures and pre-season chlorophyll-*a* show significant skill in many of the selected lakes. For these lakes, season-ahead prediction of algae metrics may provide actionable information to lake managers and public health officials based on easily accessible gridded datasets and basic water quality monitoring.

5. Conclusions

In this paper, season ahead predictions for July-October algae *magnitude* and *duration* are developed for 135 lakes identified across the Northeast and Midwest U.S. to inform lake management decisions prior to peak algae biomass. Prediction models are conditioned on local and

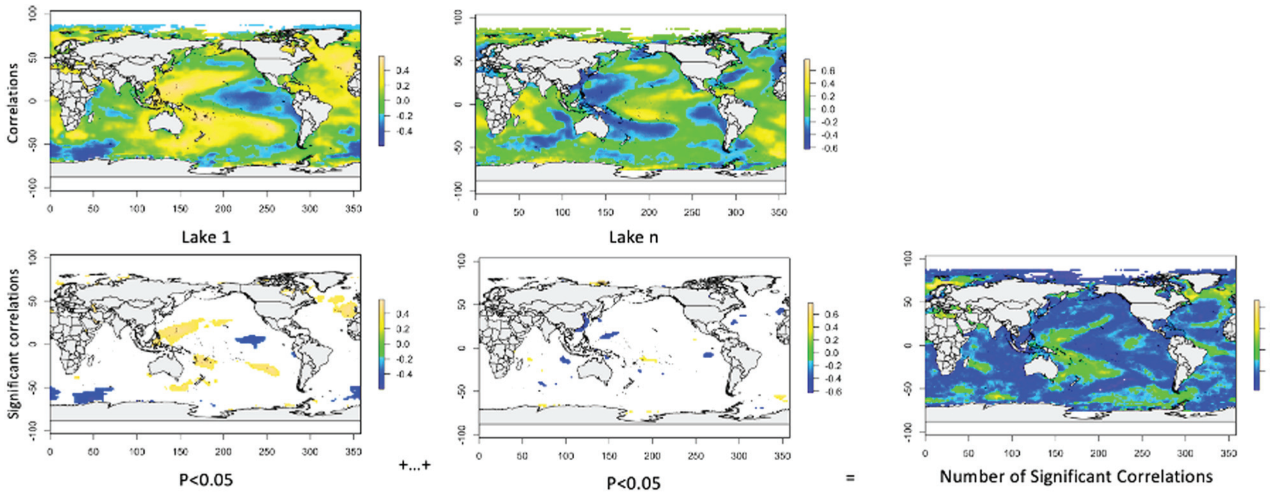

Fig. A.2. Sea surface temperature grid selection process; example from the Northeast region.

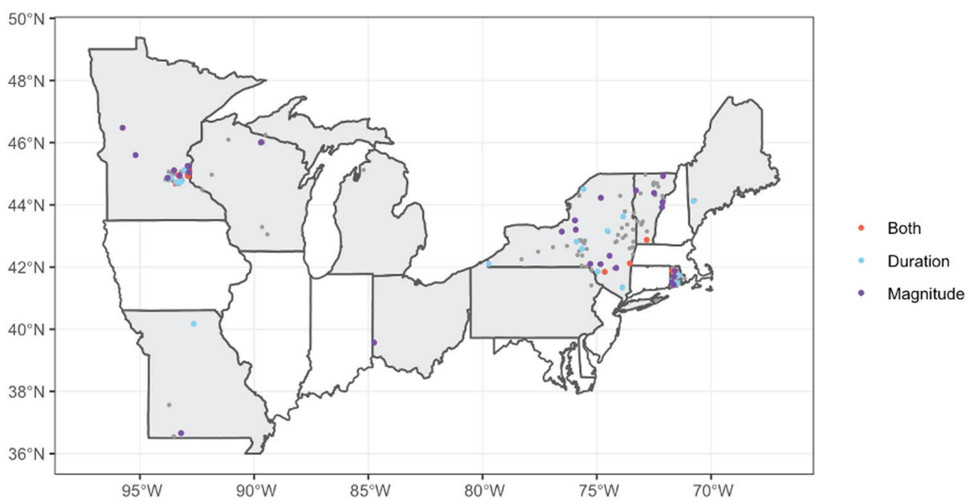

Fig. A.3. Lakes for which at least one hydroclimatic predictor is statistically significantly correlated with *magnitude, duration*, or both. Other lakes are shown in gray.

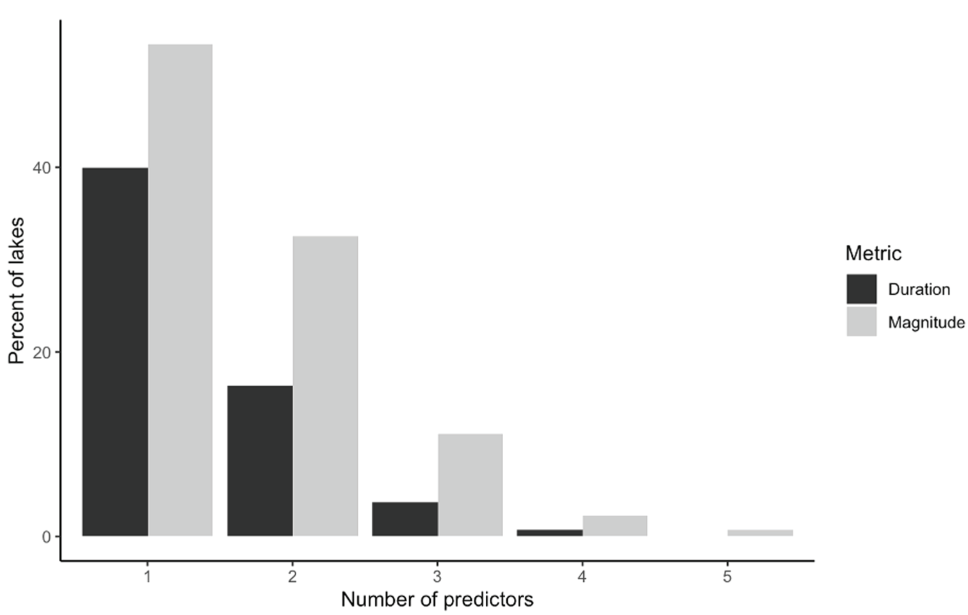

Fig. A.4. The number of predictors retained for each *magnitude* and *duration* lake model.

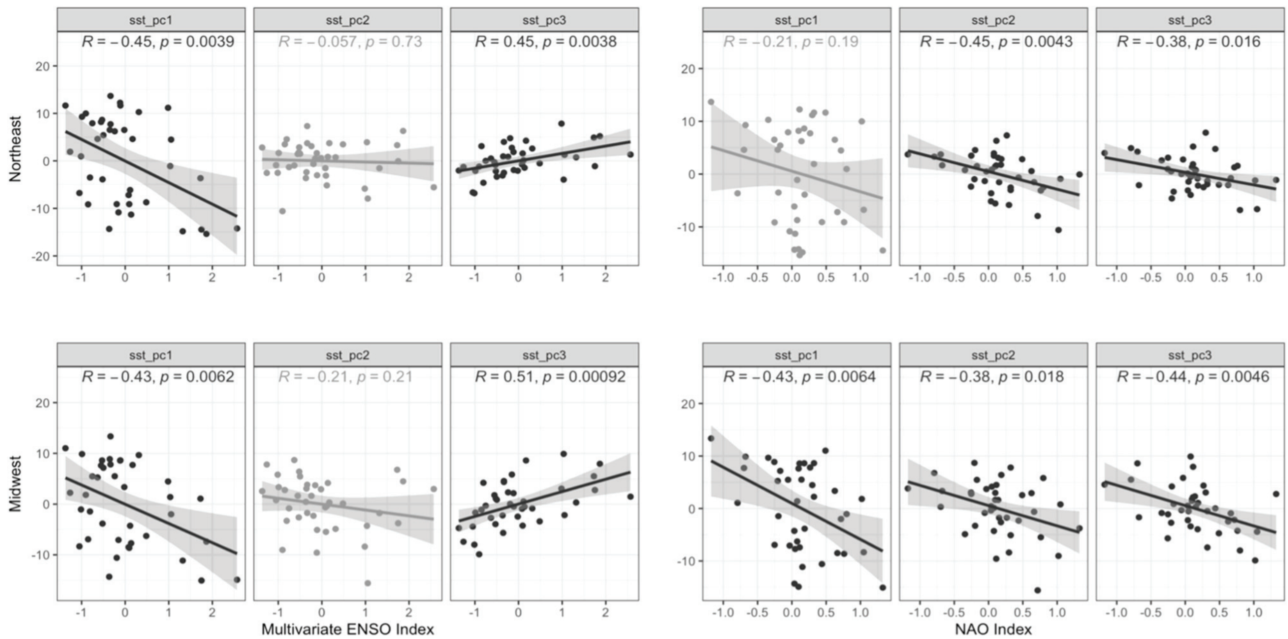

Fig. A.5. Correlation between selected SST principal components and NAO and MEI indices for the Northeast and Midwest U.S.

global scale pre-season (March-June), readily available, gridded hydroclimatic variables and pre-season chlorophyll-*a*. Global SST and pre-season chlorophyll-*a* are the most common sources of predictive power across lake models. SST grids selected for prediction model development are concentrated in the northern Atlantic and equatorial Pacific, with characteristics of both ENSO and NAO.

Forecasting models outperform climatology in 87% (70%) of *magnitude* models and in 96% (94%) of *duration* models based on HSS (RPSS). Additionally, skillful prediction of elevated algae metrics, based on *magnitude* and *duration*, is evident in more than half of the included lakes. As cultural eutrophication fuels an expansion of harmful algae in lakes across the U.S., prediction tools to inform water quality management, particularly those conditioned on easily accessible data, may incentive preparedness actions and lake management decisions toward protecting public health and informing recreational activities.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Paul Block reports financial support was provided by National Science Foundation.

Data Availability

The model code used in this study is available at https://github.com/mrwbeal/LAGOSForecast. All other data is public available or available on request.

Acknowledgements

This work is supported by an NSF CAREER project grant (1845783) and the UW-Madison graduate school. We also acknowledge the LAGOS Research Program and North Temperate Lakes Long Term Ecological Research Program for providing water quality data. The model code used in this study is available at https://github.com/mrwbeal/LAGOSForecast.

Appendix

References

Alexander, S., Wu, S., Block, P., 2019. Model Selection Based on Sectoral Application Scale for Increased Value of Hydroclimate-Prediction Information. J. Water Resour. Plann. Manage. 145, 04019006 https://doi.org/10.1061/(ASCE)WR.1943-5452.0001044 v.

Angradi, T., Ringold, P., Hall, K., 2018. Water clarity measures as indicators of recreational benefits provided by US lakes: Swimming and aesthetics. Ecol. Indic. 93, 10051-1019. https://doi.org/10.1016/j.ecolind.2018.06.001.

Arhonditsis, G.B., Winder, M., Brett, M.T., Schindler, D.E., 2004. Patterns and mechanisms of phytoplankton variability in Lake Washington (USA). Water Res. 38, 4013–4027 v.

Barnett, T.P., 1981. Statistical prediction of North American air temperatures from Pacific predictors. Mon. Weather Rev. 109, 1021–1041 v.

Barnston, A.G., 1994. Linear statistical short-term climate predictive skill in the Northern Hemisphere. J. Clim. (7), 1513–1564. https://doi.org/10.1175/1520-0442(1994)007 v.

Beal, M.R.W., O'Reilly, B., Hietpas, K.R., Block, P., 2021. Development of a sub-seasonal cyanobacteria prediction model by leveraging local and global scale predictors. Harmful Algae 108, 102100 v.

Brooks, B.W., Lazorchak, J.M., Howard, M.D.A., Johnson, M.V., Morton, S.L., Perkins, D. A.K., Reavie, E.D., Scott, G.I., Smith, S.A., Steevens, J.A., 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? Environ. Toxicol. Chem. 35, 6–13 v.

Carlson, R.E., 1977. A trophic state index for lakes 1. Limnol. Oceanogr. 22, 361–369 v.

Carmichael, W.W., 2001. Health effects of toxin-producing cyanobacteria:"The CyanoHABs": human and ecological risk assessment. Int. J. (7), 1393–1407 v.

Carmichael, W.W., Boyer, G.L., 2016. Health impacts from cyanobacteria harmful algae blooms: implications for the North American Great Lakes. Harmful Algae 54, 194–212. https://doi.org/10.1016/j.hal.2016.02.002 v.

Carpenter, S.R., Booth, E.G., Kucharik, C.J., 2018. Extreme precipitation and phosphorus loads from two agricultural watersheds. Limnol. Oceanogr. 63, 1221–1233 v.

Carpenter, S.R., Booth, E.G., Kucharik, C.J., Lathrop, R.C., 2015. Extreme daily loads: role in annual phosphorus input to a north temperate lake. Aquat. Sci. 77, 71–79 v.

Carpenter, S.R., Kitchell, J.F., Hodgson, J.R., Cochran, P.A., Elser, J.J., Elser, M.M., Lodge, D.M., Kretchmer, D., He, X., von Ende, Cn, 1987. Regulation of lake primary productivity by food web structure. Ecology 68, 1863–1876 v.

Castillo, M.M., Allan, J.D., Brunzell, S., 2000. Nutrient Concentrations and Discharges in a Midwestern agricultural Catchment. Wiley Online Library.

Chen, Q., Guan, T., Yun, L., Li, R., Recknagel, F., 2015. Online forecasting chlorophyll a concentrations by an auto-regressive integrated moving average model. Feasibilities and potentials: Harmful Algae 43, 58–65 v.

Cho, J., Shin, C.-.M., Choi, H.-.K., Kim, K.-.H., Choi, J.-.Y., 2016. Development of an integrated method for long-term water quality prediction using seasonal climate forecast. Proc. Int. Assoc. Hydrol. Sci. 374, 175–185 v.

Cottingham, K.L., Rusak, J.A., Leavitt, P.R., 2000. Increased ecosystem variability and reduced predictability following fertilisation: evidence from palaeolimnology. Ecol. Lett. 3, 340–348 v.

Dalton, C., 2021. Kaptur Announces $1.77 Million for Great Lakes Harmful Algal Bloom Research Projects. https://s3.documentcloud.org/documents/21094180/press-release-kaptur-announces-17m-for-great-lakes-harmful-algal-bloom-research-projects.pdf.

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T., Thornbrugh, D.J., 2009. Eutrophication of US freshwaters: Analysis of Potential Economic Damages.

Elliott, J.A., Jones, I.D., Thackeray, S.J., 2006. Testing the sensitivity of phytoplankton communities to changes in water temperature and nutrient load, in a temperate lake. Hydrobiologia 559. https://doi.org/10.1007/s10750-005-1233-y v.

Enfield, D.B., Mestas-Nuñez, A.M., Trimble, P.J., 2001. The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophys. Res. Lett. 28, 2077–2080 v.

Eppley, R.W., 1972. Temperature and phytoplankton growth in the sea. Fish. Bull. 70, 1063–1085 v.

Epstein, E.S., 1969. A Scoring System for Probability Forecasts of Ranked Categories. J. Appl. Meteorol. Climatol. (8), 985–987. https://doi.org/10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2.

Fee, E.J., Hecky, R.E., Regehr, G.W., Hendzel, L.L., Wilkinson, P., 1994. Effects of lake size on nutrient availability in the mixed layer during summer stratification. Can. J. Fish. Aquat.Sci. 51, 2756–2768 v.

Giannini, A., Kushnir, Y., Cane, M.A., 2000. Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. J Clim (13), 297–311 v.

Glavan, M., Ceglar, A., Pintar, M., 2015. Assessing the impacts of climate change on water quantity and quality modelling in small Slovenian Mediterranean catchment–lesson for policy and decision makers. Hydrol. Process 29, 3124–3144 v.

Glibert, P.M., Burkholder, J.M., 2006. The complex relationships between increases in fertilization of the earth, coastal eutrophication and proliferation of harmful algal blooms. Ecology of Harmful Algae. Springer, pp. 341–354.

Haygarth, P.M., Jarvis, S.C., 1997. Soil derived phosphorus in surface runoff from grazed grassland lysimeters. Water Res. 31, 140–148. https://doi.org/10.1016/S0043-1354(99)80002-5 v.

Heidke, P., 1926. Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst. Geogr. Ann. 8, 301–349 v.

Helsel, D.R., Hirsch, R.M., 1992. Statistical Methods in Water Resources. Elsevier, p. 49 v.

Ho, J.C., Michalak, A.M., Pahlevan, N., 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. Nature 574, 667–670 v.

Houlahan, J.E., McKinney, S.T., Anderson, T.M., McGill, B.J., 2017. The priority of prediction in ecological understanding. Oikos 126, 1–7 v.

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. 16, 471–483 v.

International Joint Commission, 2014. A Balanced Diet For Lake Erie.

Kane, D.D., Conroy, J.D., Peter Richards, R., Baker, D.B., Culver, D.A., 2014. *Re*-eutrophication of Lake Erie: correlations between tributary nutrient loads and phytoplankton biomass. J. Great Lakes Res. 40, 496–501. https://doi.org/10.1016/j.jglr.2014.04.004 v.

Kleinman, P.J.A., Srinivasan, M.S., Dell, C.J., Schmidt, J.P., Sharpley, A.N., Bryant, R.B., 2006. Role of Rainfall Intensity and Hydrology in Nutrient Transport Via Surface Runoff.

Konopka, A., Brock, T.D., 1978. Effect of temperature on blue-green algae (cyanobacteria) in Lake Mendota. Appl. Environ. Microbiol. 36, 572–576 v.

Lampert, W., Fleckner, W., Rai, H., Taylor, B.E., 1986. Phytoplankton control by grazing zooplankton: a study on the spring clear-water phase 1. Limnol. Oceanogr. 31, 478–490 v.

Lathrop, R.C., Carpenter, S.R., Stow, C.A., Soranno, P.A., Panuska, J.C., 1998. Phosphorus loading reductions needed to control blue-green algal blooms in Lake Mendota. Can. J. Fish. Aquat.Sci. 55, 1169–1178. https://doi.org/10.1139/f97-317 v.

León-Muñoz, J., Urbina, M.A., Garreaud, R., Iriarte, J.L., 2018. Hydroclimatic conditions trigger record harmful algal bloom in western Patagonia (summer 2016). Sci. Rep. 8, 1–10 v.

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

Liu, R., Wang, J., Shi, J., Chen, Y., Sun, C., Zhang, P., Shen, Z., 2014. Runoff characteristics and nutrient loss mechanism from plain farmland under simulated rainfall conditions. Sci. Total Environ. 468, 1069–1077 v.

Mander, Ü., Kull, A., Kuusemets, V., Tamm, T., 2000. Nutrient runoff dynamics in a rural catchment: influence of land-use changes, climatic fluctuations and ecotechnological measures. Ecol. Eng. 14, 405–417. https://doi.org/10.1016/S0925-8574(99)00064-6 v.

Markowski, G.R., North, G.R., 2003. Climatic influence of sea surface temperature: evidence of substantial precipitation correlation and predictability. J. Hydrometeorol. 4, 856–877. https://doi.org/10.1175/1525-7541(2003)004<0856:CIOSST>2.0.CO;2 v.

Me, W., Hamilton, D.P., McBride, C.G., Abell, J.M., Hicks, B.J., 2018. Modelling hydrology and water quality in a mixed land use catchment and eutrophic lake: effects of nutrient load reductions and climate change. Environ. Model. Softw. 109, 114–133 v.

Moss, B., 2012. Cogs in the endless machine: lakes, climate change and nutrient cycles: a review. Sci. Total Environ. 434, 130–142 v.

O'Neil, J.M., Davis, T.W., Burford, M.A., Gobler, C.J., 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. Harmful Algae 14, 313–334 v.

Paerl, H.W., Paul, V.J., 2012. Climate change: links to global expansion of harmful cyanobacteria. Water Res. 46, 1349–1363 v.

Park, J.-Y., Stock, C.A., Dunne, J.P., Yang, X., Rosati, A., 2019. Seasonal to multiannual marine ecosystem prediction with a global Earth system model. Science 365, 284–288 v.

Patel, J., Parshina-Kottas, Y., 2017. Miles of Algae Covering Lake Erie. New York Times. https://www.nytimes.com/interactive/2017/10/03/science/earth/lake-erie.html.

Qian, S.S., Stow, C.A., Rowland, F.E., Liu, Q., Rowe, M.D., Anderson, E.J., Stumpf, R.P., Johengen, T.H., 2021. Chlorophyll a as an indicator of microcystin: short-term forecasting and risk assessment in Lake Erie. Ecol. Indic. 130, 108055 v.

Reutter, J.M., Ciborowski, J., DePinto, J., Bade, D., Baker, D., Bridgeman, T.B., Culver, D. A., Davis, S., Dayton, E., Kane, D., 2011. Lake Erie nutrient Loading and Harmful Algal blooms: Research Findings and Management Implications. Lake Erie Millennium Network Synthesis Team. Final Report of the.

Reynolds, C.S., 1984. The Ecology of Freshwater Phytoplankton. Cambridge university press.

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

Robarts, R.D., Zohary, T., 1987. Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. N. Z. J. Mar. Freshwater Res. 21, 391–399 v.

Roelke, D., Buyukates, Y., 2001. The diversity of harmful algal bloom-triggering mechanisms and the complexity of bloom initiation: human and Ecological Risk Assessment. Int. J. 7, 1347–1362 v.

Ropelewski, C.F., Halpert, M.S., 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. Mon. Weather Rev. 115, 1606–1626 v.

da Rosa Wieliczko, A., Crossetti, L.O., Cavalcanti, J.R., Hessel, M.S., da Motta-Marques, D., Rodrigues, L.R., 2021. Meteorological drivers and ENSO influence on phytoplankton biomass dynamics in a shallow subtropical lake. Environ. Monit. Assess. 193, 536. https://doi.org/10.1007/s10661-021-09288-4 v.

Rousseaux, C.S., Gregg, W.W., Ott, L., 2021. Assessing the Skills of a Seasonal Forecast of Chlorophyll in the Global Pelagic Oceans. Remote Sens. (Basel) 13, 1051 v.

Royer, T.V., David, M.B., Gentry, L.E., 2006. Timing of riverine export of nitrate and phosphorus from agricultural watersheds in Illinois: implications for reducing nutrient loading to the Mississippi River. Environ. Sci. Technol. 40, 4126–4131. https://doi.org/10.1021/es052573n v.

Rusak, J.A., Tanentzap, A.J., Klug, J.L., Rose, K.C., Hendricks, S.P., Jennings, E., Laas, A., Pierson, D., Ryder, E., Smyth, R.L., 2018. Wind and trophic status explain within and among-lake variability of algal biomass. Limnol. Oceanogr. Lett. 3, 409–418 v.

Salmaso, N., Buzzi, F., Garibaldi, L., Morabito, G., Simona, M., 2012. Effects of nutrient availability and temperature on phytoplankton development: a case study from large lakes south of the Alps. Aquat. Sci. 74, 555–570 v.

Schindler, D.W., 1971. Carbon, nitrogen, and phosphorus and the eutrophication of freshwater lakes 1. J. Phycol. 7, 321–329 v.

Schindler, D.W., 1977. Evolution of phosphorus limitation in lakes: natural mechanisms compensate for deficiencies of nitrogen and carbon in eutrophied lakes. Science 195, 260–262 v.

Schindler, D.W., 1978. Factors regulating phytoplankton production and standing crop in the world's freshwaters. Limnol. Oceanogr. 23, 478–486 v.

Scordo, F., Lottig, N.R., Fiorenza, J.E., Culpepper, J., Simmons, J., Seitz, C., Krynak, E.M., Suenaga, E., Chandra, S., 2022. Hydroclimate variability affects habitat-specific (open water and littoral) lake metabolism. Water Resour. Res., e2021WR031094

Shuter, B.J., Schlesinger, D.A., Zimmerman, A.P., 1983. Empirical predictors of annual surface water temperature cycles in North American lakes. Can. J. Fish. Aquat.Sci. 40, 1838–1845 v.

Singh, S.P., Singh, P., 2015. Effect of temperature and light on the growth of algae species: a review. Renewable Sustainable Energy Rev. 50, 431–444 v.

Sinha, E., Michalak, M., Balaji, V., 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. Science 357, 405–408. https://doi.org/10.1126/science.aan2409 v.

Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environ. Sci. Pollut. Res. Int. 10, 126–139. https://doi.org/10.1065/espr2002.12.142 v.

Soranno, P.A., et al., 2017. LAGOS-NE: a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes. Gigascience 6. https://doi.org/10.1093/gigascience/gix101 v.

Stow, C.A., Carpenter, S.R., Lathrop, B.C., 1997. A Bayesian observation error model to predict cyanobacterial biovolume from spring total phosphorus in Lake Mendota. Wisconsin: Can. J. Fisheries Aquatic Sci. 54, 464–473. https://doi.org/10.1139/f96-279 v.

Stumpf, R.P., Johnson, L.T., Wynne, T.T., Baker, D.B., 2016. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. J. Great Lakes Res. 42, 1174–1183. https://doi.org/10.1016/j.jglr.2016.08.006 v.

Sunda, W.G., Graneli, E., Gobler, C.J., 2006. Positive feedback and the development and persistence of ecosystem disruptive algal blooms 1. J. Phycol. 42, 963–974 v.

Tootle, G.A., Piechota, T.C., Singh, A., 2005. Coupled oceanic-atmospheric variability and US streamflow. Water Resour. Res. 41 v.

Trombetta, T., Vidussi, F., Roques, C., Mas, S., Scotti, M., Mostajir, B., 2021. Co-occurrence networks reveal the central role of temperature in structuring the plankton community of the Thau Lagoon. Sci. Rep. 11, 1–14 v.

Vanni, M.J., Temte, J., 1990. Seasonal patterns of grazing and nutrient limitation of phytoplankton in a eutrophic lake. Limnol. Oceanogr. 35, 697–709 v.

Visbeck, M.H., Hurrell, J.W., Polvani, L., Cullen, H.M., 2001. The North Atlantic Oscillation: past, present, and future. Proc. Natl. Acad. Sci. 98, 12876–12877 v.

Vollenweider, R.A., 1971. Scientific Fundamentals of the Eutrophication of Lakes and Flowing waters, With Particular Reference to Nitrogen and Phosphorus As Factors in Eutrophication. Organisation for economic co-operation and development, Paris.

Wilkinson, G.M., Walter, J.A., Buelo, C.D., Pace, M.L., 2021. No evidence of widespread algal bloom intensification in hundreds of lakes. Front. Ecol. Environ. https://doi.org/10.1002/fee.2421 vn/a.

Wines, M., 2014. Behind Toledo's Water Crisis, a Long-Troubled Lake Erie. New York Times. https://www.nytimes.com/2014/08/05/us/lifting-ban-toledo-says-its-water-is-safe-to-drink-again.html.

Woolway, R.I., Kraemer, B.M., Lenters, J.D., Merchant, C.J., O'Reilly, C.M., Sharma, S., 2020. Global lake responses to climate change. Nature Rev. Earth Environ. v 1, 388–403.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Fahnenstiel, G.L., Dyble, J., Schwab, D.J., Joshi, S.J., 2013. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: development and initial evaluation. J. Great Lakes Res. 39, 90–99 v.

Xiao, X., He, J., Yu, Y., Cazelles, B., Li, M., Jiang, Q., Xu, C., 2019. Teleconnection between phytoplankton dynamics in north temperate lakes and global climatic oscillation by time-frequency analysis. Water Res. 154, 267–276. https://doi.org/10.1016/j.watres.2019.01.056 v.

Zhang, H., Hu, W., Gu, K., Li, Q., Zheng, D., Zhai, S., 2013. An improved ecological model and software for short-term algal bloom forecasting. Environ. Model. Softw. 48, 152–162 v.

Zhu, S., Ptak, M., Yaseen, Z.M., Dai, J., Sivakumar, B., 2020. Forecasting surface water temperature in lakes: a comparison of approaches. J. Hydrol. (Amst.) 585, 124809 v.