

Original Article

Development of a sub-seasonal cyanobacteria prediction model by leveraging local and global scale predictors

Maxwell R.W. Beal^{*}, Bryan O'Reilly, Kaitlynn R. Hietpas, Paul Block

Department of Civil and Environmental Engineering, University of Wisconsin, Madison, WI, USA

ARTICLE INFO

Keywords:

Cyanobacteria
Sub-seasonal forecasting
Statistical prediction
Water resources management
Water quality

ABSTRACT

In recent decades, cultural eutrophication of coastal waters and inland lakes around the world has contributed to a rapid expansion of potentially toxic cyanobacteria, threatening aquatic and human systems. For many locations, a complex array of physical, chemical, and biological variables leads to significant inter-annual variability of cyanobacteria biomass, modulated by local and large-scale climate phenomena. Currently, however, minimal information regarding expected summertime cyanobacteria biomass conditions is available prior to the season, limiting proactive management and preparedness strategies for lake and beach safety. To address this, sub-seasonal (two-month) cyanobacteria biomass prediction models are developed, drawing on pre-season predictors including stream discharge, phosphorus loads, a floating algae index, and large-scale sea-surface temperature regions, with an application to Lake Mendota in Wisconsin. A two-phase statistical modeling approach is adopted to reflect identified asymmetric relationships between predictors (drivers of inter-annual variability) and cyanobacteria biomass levels. The model illustrates promising performance overall, with particular skill in predicting above normal cyanobacteria biomass conditions which are of primary importance to lake and beach managers.

1. Introduction

Cyanobacteria are aquatic photosynthetic microorganisms that can produce a range of toxins (Carmichael et al., 1994; Elliot, 2012) and potentially lead to harmful algal blooms with adverse impacts on ecosystems, economies, and human and animal health worldwide (Paerl et al., 2001; Guo, 2007; Carmichael and Boyer, 2016). Proliferation of cyanobacteria can negatively impact water quality by increasing turbidity and depleting oxygen through microbial degradation, leading to the death of aquatic macrophytes, fish, and invertebrates (Rabalais et al., 2010; Huisman et al., 2018). Additionally, toxins produced by cyanobacteria, or cyanotoxins, can result in mortality of animals, birds, and fish, weakening community structure (Paerl et al., 2001). Cyanotoxins are also responsible for human health issues ranging from contact irritation to lethal poisoning (Carmichael and Boyer, 2016). Since 2009, the Center for Disease Control National Outbreak Reporting System has documented eight outbreaks and 331 illnesses related to cyanotoxin exposure in the U.S. (221 - recreational exposure, 110 - drinking water) (Centers for Disease Control and Prevention (CDC) 2020). Cyanobacteria-related impacts on environmental and human health lead

to economic stress as well. A study conducted in Australia assessed the cost of environmental management attributable to cyanobacteria, including wastewater treatment and rehabilitation of water resources, to range from \$180 million to \$240 million annually (Atech, 2000). More broadly, potential losses related to cultural eutrophication of freshwater in the U.S., a known driver of cyanobacteria biomass, total over \$2.2 billion annually (Dodds et al., 2009).

In recent years, concerns regarding the global proliferation of cyanobacteria-related to eutrophication and climate change have grown (Paerl and Huisman, 2008; Huisman et al., 2018). Analysis of cyanobacteria pigment in north temperate and subarctic sediment cores has revealed an acceleration of cyanobacteria growth since 1945, and a disproportionate increase in cyanobacteria relative to other, less harmful phytoplankton (Taranu et al., 2015). In the U.S., cyanobacteria blooms have received notable attention in larger water bodies. For example, a 2014 bloom in Lake Erie drew national attention after causing a shutdown of the drinking water supply in Toledo, Ohio, due to dangerous levels of cyanotoxins (Patel and Parshina-kottas, 2017). Other waterbodies such as Lake Okeechobee in Florida, and the Chesapeake Bay have been the focus of cyanobacteria-related research and

^{*} Corresponding author.

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

<https://doi.org/10.1016/j.hal.2021.102100>

Received 29 April 2021; Received in revised form 21 August 2021; Accepted 29 August 2021

Available online 9 September 2021

1568-9883/© 2021 Elsevier B.V. All rights reserved.

legislation for the past 20 years (University of Maryland Center for Environmental Science 2015; Stern and Kornfield, 2016). A globally robust relationship between algal biomass and nutrient input (Smith, 2003) suggests that small eutrophic lakes are significantly impacted as well, but have received less attention.

Typically, cyanobacteria biomass peaks during the summer season (June-Aug or Sept) in the northern hemisphere, and is determined by the interactions of many physical, chemical, and biological variables throughout the spring and summer (Robarts and Zohary, 1987; Paerl, 1988; Stow et al., 1997; Carpenter et al., 2018). Numerous studies have shown that nitrogen and phosphorus are key nutrients for cyanobacteria growth, however, phosphorus is generally considered the limiting nutrient for algal growth in freshwater systems (Schindler 1977; Downing et al., 2001; Smith et al., 2003; Xu et al., 2010). Agricultural use of nitrogen and phosphorus fertilizers creates nutrient-rich runoff that spurs productivity in lakes (Hart et al., 2004; Kronvang et al., 2005). Phosphorus is also known to accumulate in sediments which settle in lakes and subsequently release stored phosphorus into the water column (Søndergaard et al., 2003). Ratios of total nitrogen to total phosphorus have also been shown to have some influence on cyanobacteria biomass, however, the relationship may be more complex (Håkanson et al., 2007). Physical lake characteristics, including temperature and stratification strength, also regulate cyanobacteria biomass. The photosynthetic capacities of many cyanobacteria species are maximized at high temperatures (Konopka and Brock, 1978; Robarts and Zohary, 1987), and gas vacuoles, used to regulate buoyancy, provide cyanobacteria with a competitive advantage over other phytoplankton groups in stratified water columns (Reynolds et al., 1987; Walsby et al., 1997; Huber et al., 2012). Additionally, cyanobacteria are known to have a high physiological tolerance, allowing them to grow in temperatures unfavorable to other organisms (Briand et al., 2004; Pearl and Huisman, 2008). Many of the variables that may influence cyanobacteria growth are regulated by local climatic conditions which are, in turn, influenced by large-scale climate phenomena through atmospheric teleconnections. These variables can affect long-term trends in cyanobacteria growth but also lead to significant inter-annual variability in cyanobacteria biomass.

To address inter-annual variations in cyanobacteria biomass – most notably when levels are elevated – lake managers may benefit from advanced indication of expected conditions, particularly if there are explicit actions they can take. Most efforts in this vein focus on short-term (days to weeks) cyanobacteria forecasts with the goal of providing early warnings of elevated levels or bloom formation and growth (Wynne et al., 2013; Zhang et al., 2015). These forecasts allow public health officials and lake managers to respond to immediate threats. For example, Zhang et al. (2015) list emergency responses to elevated cyanobacteria in Lake Taihu, China, to include water diversion for flushing, algal bloom collection and removal, and emergency measures for securing drinking water. On Lake Erie, Wynne et al. (2013) suggest other short-term actions such as alleviating taste and odor issues in drinking water and posting warning signs at beaches. In contrast, longer-lead (months to seasons) forecasts of elevated cyanobacteria biomass may provide opportunities for proactive lake management not possible at short timescales. Lake and beach managers may be able to adjust budgets for water quality testing, augment lifeguard training, and increase public awareness. These longer-lead forecasts are not intended to replace short-term forecasts, but ideally work in concert to provide managers a suite of actions at various lead times.

Considerable progress has been made in the development of season-ahead forecasts to address water quantity management (e.g. Chiew et al., 2003; Delorit et al., 2017; Giuliani et al. 2019a; Baker et al., 2019), however significantly less focus has been devoted to the application of season-ahead forecasts for water quality management. Towler et al. (2010) applied seasonal forecasts to turbidity thresholds for drinking water in the Northwest U.S. with moderate success. On a sub-seasonal scale, Wang et al. (2013) developed a bootstrapped wavelet neural

network to forecast monthly ammonia nitrogen and dissolved oxygen in the Harbin region of China. Recknagel et al. (2017) developed a microcystin cell concentration forecast for a similar timescale, effective for forecast horizons of up to 30 days. Previous research in cyanobacteria forecasting, however, has predominantly focused on exploring alternative modeling techniques, relying primarily on spring phosphorus loads, with minimal consideration of how season-ahead large-scale and local hydroclimatic drivers may additionally inform variations in summertime cyanobacteria biomass (Stow et al., 1997; Lathrop et al., 1998; Stumpf et al., 2016). Recently, the potential for developing sub-seasonal (i.e. within-season) forecasts for application to water management has garnered attention (Vitart et al., 2012; Vitart et al., 2018) with the intention that such forecasts could bridge the gap between seasonal and short-term time scales (Shentsis and Ben-Zvi, 1999; Vitart, 2014). In conversations with lake and public health managers, there is an expressed desire to understand how cyanobacteria abundance may be changing throughout the summer season, and if a prediction update is possible. A sub-seasonal forecast of cyanobacteria biomass may indicate if expected cyanobacteria conditions are shifting within the season, providing managers with an opportunity to change the frequency of water quality monitoring, public engagement strategies, and prepare emergency resources for recreators and drinking water facilities before the potential for cyanobacteria productivity peaks. This study investigates relevant pre- and within-season local and global scale drivers of inter-annual variability in summertime cyanobacteria biomass and presents the development and verification of a sub-seasonal forecasting framework for cyanobacteria conditions. Finally, this study explores how a sub-seasonal forecast may be effectively paired with a seasonal forecast for holistic lake management.

1.1. Case study

Lake Mendota, in Madison, Wisconsin, is often labeled as one of the most studied lakes in the world (Brezonik et al., 1968; Brock, 1985; Aoki, 1989; Lathrop, 2007) (Figure 1). Mendota's 596-square kilometer watershed is highly urbanized (21%) and agricultural (53%) (Genskow et al., 2012). Municipal wastewater discharge fueled eutrophication in Lake Mendota from the 1940's-1970's, however, in recent years urban and agricultural development in the Mendota watershed has maintained the state of high productivity in the lake (Lathrop et al., 1998; Lathrop, 2007). Cyanobacteria blooms have occurred in Lake Mendota since the late 1800's but have become a more serious concern as cultural eutrophication has progressed over the last century, making blooms a common summertime phenomenon (Brock, 1985; Lathrop and Carpenter, 1992; Lathrop et al., 1998).

Seasonal forecasts have been produced for summertime cyanobacteria biomass on Lake Mendota since 2015 (Soley, 2016). The authors use a principal component analysis and regression modelling approach conditioned on season-ahead (March-May) local and global scale predictors to generate probabilistic forecasts of average cyanobacteria biomass for June-August (released on June 1). The highest cyanobacteria biomass concentrations, however, have historically occurred in July and August (Fig. 2), further motivating the potential utility of a sub-seasonal forecast by updating later in the season.

Numerous large-scale climate phenomena influence climate conditions in the upper Midwestern US, however one of the most prominent teleconnection patterns affecting precipitation and temperature is the El Niño Southern Oscillation (ENSO) (Ropelewski and Halpert, 1986; 1987). ENSO's influence globally is widely studied and generally understood, acting independently or interacting with other large-scale climate phenomena such as the Pacific Decadal Oscillation (Kahya and Dracup 1993; Shabbar and Skinner, 2004). The state of these climate phenomena and the atmospheric-oceanic system is an important factor in local climate variability, and therefore regulate many processes important to cyanobacteria growth (Justić et al., 2005; Zhang et al., 2012). Two studies of lake ice cover records in Wisconsin lakes found

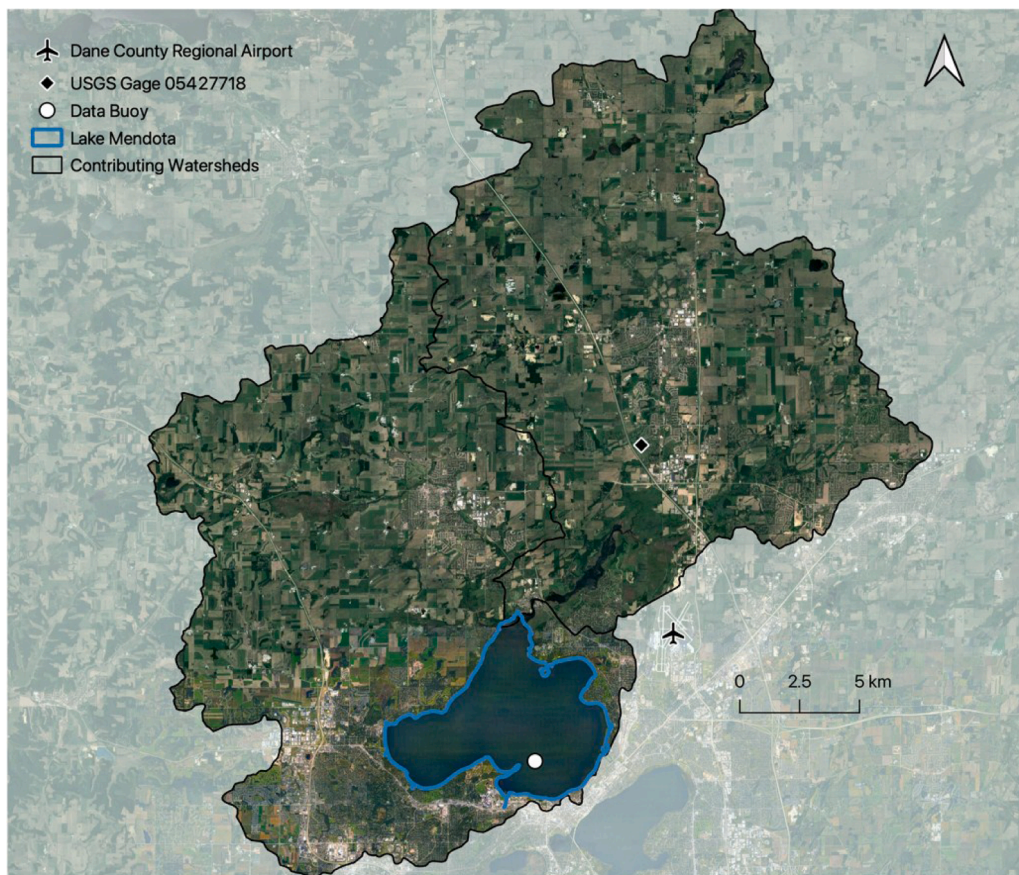


Fig. 1. Map of Lake Mendota, contributing watersheds, and sources of local data.

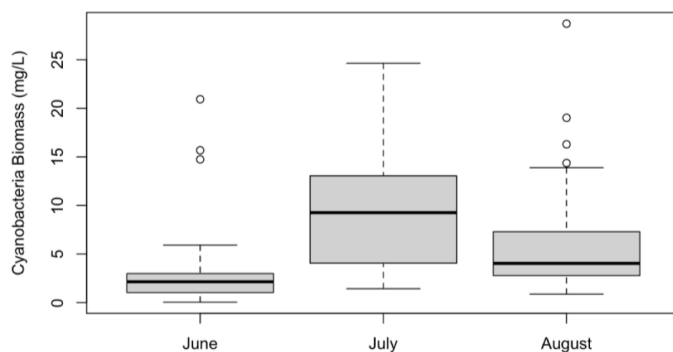


Fig. 2. June, July, and August average cyanobacteria biomass at the LTER buoy in Lake Mendota (1995–2017).

that warmer air temperatures lead to early spring ice breakup in strong El Niño years (Anderson et al., 1996; Robertson et al., 2000). Additionally, an analysis of winter precipitation in the Great Lakes region found evidence for an asymmetric response to measures of ENSO, with more distinct linkages to La Niña conditions (Fu and Steinschneider, 2019).

Lake Mendota is included in the National Science Foundation's North Temperate Lakes - Long-Term Ecological Research (LTER) network and has a wealth of high-quality, long-term ecological data (LTER 2020). Access to cyanobacteria data in addition to extensive ecological, hydrological, and climatological datasets make Lake Mendota an ideal case-study for the application of seasonal (June-August) and sub-seasonal (July-August) cyanobacteria forecasts. Cyanobacteria biomass data has been collected at the LTER buoy in Lake Mendota each summer

since 1995 (Magnuson et al., 2020a). Composite samples for 0-8 meters of depth are typically collected 2-4 times a month from June through August.

2. Methods

2.1. Drivers of variability in cyanobacteria biomass

In Lake Mendota, cyanobacteria biomass is generally most apparent from June-August (Fig. 2), with elevated levels typically observable in July and August. The existing seasonal prediction model issues a June-August average cyanobacteria biomass forecast at the beginning of June, whereas the sub-seasonal July-August prediction model proposed here focuses on the peak months, taking advantage of June observations, and issues a forecast of average July-August cyanobacteria biomass at the beginning of July.

Numerous drivers and factors at local to global scales influence inter-annual cyanobacteria productivity. From a forecasting perspective, ideal predictors include pre-season (e.g. April-June), observable hydro-climatic and landscape variables that effect the state of the lake system into July-August. Potential predictors are identified based on previous literature regarding cyanobacteria dynamics and/or correlation analysis. Predictors that correlate with July-August average cyanobacteria biomass at the 95% confidence level ($P < 0.05$) are considered statistically significant and added to the suite of potential predictors.

As discussed previously, phosphorus is a well-established driver of cyanobacteria biomass. Strong correlations between phosphorus in contributing waters and cyanobacteria biomass has been demonstrated repeatedly (Smith, 1985; Stow et al., 1997; Lathrop et al., 1998; Downing et al., 2001; Håkanson et al., 2007). Phosphorus data for the Lake Mendota case study are extracted from USGS station 05427718

located on the Yahara River at Windsor, WI for the month of June (USGS, 2020). Relatedly, local scale hydroclimatic variables influencing transport of phosphorus across the landscape prior to July are also important drivers of external phosphorous loading. These include extreme precipitation events, discharge, soil moisture, and suspended sediments (Michalak, 2016; Motew et al., 2017; Carpenter et al., 2018). Precipitation events can wash high concentrations of nutrients off the landscape and into surface waters, contributing to eutrophication (Sinha et al., 2017). Agricultural watersheds similar to the Lake Mendota watershed are particularly vulnerable to phosphorus loading driven by precipitation events (Garnache et al., 2016; Carpenter et al., 2018). Most of the phosphorus loading in the Mendota watershed occurs in a relatively small number of large loading events (Carpenter et al., 2015). Similarly, soil moisture conditions regulate infiltration versus direct runoff into rivers or lakes. Intense loading events occurring in June, represented by phosphorus load, discharge, and suspended sediments, have the potential to alter phosphorus availability for cyanobacteria later in the summer (Lathrop and Carpenter, 2014). Precipitation data was obtained from the Midwest Regional Climate Center for March-May (Wuertz et al., 2020). Soil moisture data in the Mendota watershed comes from the North American Land Data Assimilation System (NLDAS) for June (Mocko, 2013). Both discharge and suspended sediment loads are from USGS station 05427718 for the month of June (U.S. Geological Survey 2020a).

Variables related to in-situ productivity, including nitrate + nitrite and total unfiltered phosphorus are also considered. While phosphorus has historically been considered the primary limiting nutrient for phytoplankton in freshwater, there is evidence that inorganic nitrogen can control growth and toxicity of cyanobacteria as well (Gobler et al., 2016). These data are available in the LTER database for June (Magnuson et al., 2020b). To further assess the state of lake productivity in June, a Floating Algae Index (FAI) was generated using remotely sensed images in June from Landsat 5 (1995–1999) and MODIS (2000–2017) satellites (ORNL DAAC 2020; U.S. Geological Survey 2020b), using methods outlined by Hu (2009). The FAI has been used for mapping floating algae, including cyanobacteria, in lacustrine and coastal environments (Hu, 2009; Oyama et al., 2015). The effects of nutrient loading on algal biomass are well established, therefore, an estimation of algal biomass in June may indicate the general state of productivity in the lake and serve as a predictor of cyanobacteria productivity later in the summer (Vollenweider, 1970).

As discussed previously, elevated air temperature is thought to favor dominance of cyanobacteria through direct effects on photosynthetic capacity and indirect effects on competition. June air temperature and the number of events exceeding the 99th percentile of air temperatures for the climate reference period (1981–2010) are included in the suite of potential predictors (Gallina et al., 2011; Anneville et al., 2015). Daily temperature data from NOAA's Global Historical Climatology Network was accessed through the Midwest Regional Climate Center (Menne et al., 2012). Mean pre-season water temperature, accessed through the NTL-LTER data repository, is also evaluated as a potential predictor (Robertson, 2016; Magnuson et al., 2020c).

Variable grazing rates by *Daphnia* spp. may also influence cyanobacteria biomass. In Lake Mendota, (Lathrop et al., 1996) found that summer water clarity is significantly greater in years dominated by *D. pulicaria* compared to *D. mendotae*. This difference in clarity has been attributed to the ability of the larger-bodied *D. pulicaria* to significantly reduce summertime algal biomass, including cyanobacteria (Epp, 1996; Kasprzak and Lathrop, 1997; Sarnelle, 2007). Furthermore, there is evidence to suggest that summertime *Daphnia* biomasses are greater in Lake Mendota when *D. pulicaria* dominate in the spring months (Lathrop et al., 1999). Thus, April–May *D. pulicaria* biomass, measured at the LTER buoy, is included as a potential predictor of July–August cyanobacteria biomass (Magnuson et al., 2019).

Sea surface temperature (SST) and sea level pressure (SLP) anomalies have been well-documented to influence precipitation and temperature

on monthly to seasonal timescales by altering atmospheric flow conditions (Barnston, 1994; Giannini and Kushnir, 2000; Markowski and North, 2003; Quan et al., 2006), and are thus also considered as potential predictors. Locations of SST and SLP influencing climate conditions in the Lake Mendota watershed are further developed below. May–June SST and SLP anomalies are from NOAA's ERSST v3b and HadSLP2r datasets, respectively, and were accessed through the IRI Data Library (Allan and Ansell, 2006; Smith et al., 2008).

2.2. Prediction modeling approach

Prediction models may take a statistical/machine learning or process-based approach. Given the high complexity of limnological processes and lake dynamics, and the singular focus here on predicting cyanobacteria biomass, statistical approaches are arguably better suited (Rastetter, 2017). When sufficient observational records exist, statistical models can readily incorporate a wide variety of potential predictors at both local and global scales. This is advantageous given the vast array of variables contributing to summertime cyanobacteria growth. A principal component analysis and regression approach are selected to build the July–August cyanobacteria prediction model. Principal component analysis decomposes a multivariate dataset into orthogonal patterns (principal components, PCs) that represent the dominant signals from the original set of predictors (Block et al., 2009). This approach minimizes multi-collinearity among the predictors by construction and thus does not artificially inflate prediction skill. Here, principal components that explained more than 10% of the variance in the data are retained. Principal component regression models are fit based on the retained PCs across 1995–2017. These models take the form of Eq. 1, where α and β are the coefficients fit through ordinary least squares, PC_{1t} through PC_{nt} are the principal components retained, and Y_t is the prediction of average cyanobacteria at each time step, t .

$$Y_t = \alpha + \beta_1 * PC_{1t} + \dots \beta_n * PC_{nt} \quad (1)$$

The generalized cross-validation (GCV) score is used to select the best subset from the suite of predictors and is given as,

$$GCV = \frac{\sum_{t=1}^N \frac{e_t^2}{N}}{\left(1 - \frac{m}{N}\right)^2} \quad (2)$$

where N is the number of time steps (1995–2017), m is the number of PCs (predictors) retained in each candidate model, and e_t is the residual (difference between observed and model estimated values) at each time step, t (July–August each year). The GCV penalizes overfitting and is a good estimate of predictive risk (Craven and Wahba 1978). Using this method, the best set of predictors can be identified by evaluating a number of predictor combinations (candidate models) and selecting the combination that results in the minimum GCV score (Regonda et al., 2006). Statistical models were developed using R version 1.3.1056.

In an effort to appropriately represent teleconnection patterns between global climate phenomenon and local-scale processes that drive cyanobacteria biomass, a Nino Index Phase Analysis (Zimmerman et al. 2016; Giuliani et al. 2019a) is adopted. This method draws on the state of the atmospheric-oceanic system in months prior to the season of interest to divide a timeseries into different “mean states”. This allows for possible asymmetric relationships between “mean states” to be captured and modeled (Lee et al., 2018). Given that ENSO expresses moderate influence over climate conditions in the upper Midwestern U.S., the Multivariate ENSO Index (MEI) – consisting of SLP and SST information in the Pacific Ocean – is used to classify historical years into phases of ENSO; here two phases are adopted: positive and negative, based on MEI values averaged over May – June. Global and local-scale predictors may subsequently be evaluated for each “mean state” of the atmospheric-oceanic system represented by each phase. For the historical cyanobacteria biomass record on Lake Mendota, seven years fall into

the positive phase and 16 into the negative phase. For the years falling within each phase, regions of SST and SLP anomalies that correlated significantly with July–August cyanobacteria are selected following Zimmerman et al. (2016). Principal components of these SST and SLP regions are included as potential sub-seasonal predictors. Each of the specified predictors are evaluated independently for the positive and negative phase, as correlation with cyanobacteria biomass in one phase does not necessitate inclusion as a predictor in both phase models. Thus the important processes contributing to cyanobacteria growth in each phase of ENSO are identified, potentially leading to enhanced biomass forecasts. The Nino Index Phase Analysis was performed in Python 2.7.16 and Spyder 3.3.6 using code developed by Giuliani et al. (2019b).

To evaluate historical performance, a hindcast is undertaken, such that a year of information is dropped (drop one cross-validation), the PCs are constructed, coefficients α and β are fit based on the remaining years of data, and Y_t for the dropped year is calculated. This is repeated to create a deterministic forecast of biomass for all years. The optimal number of PCs for each model, based on the GCV, was held constant for the cross-validation in all years.

Ensemble predictions for each year in the hindcast are based on errors, defined as the difference between predicted and observed cyanobacteria biomass in the leave-one-out cross-validated approach. Errors are fit to a normal distribution, with mean zero, using a maximum likelihood estimation. For each hindcast year, 100 random draws from the distribution are added to the deterministic biomass forecast to form the ensemble (Helsel and Hirsch 1992; Zhang et al., 2016; Delroit et al., 2017; Alexander et al., 2019).

2.3. Model performance metrics

To assess model performance, observations are compared with model forecasts using five skill scores: correlation coefficients, Root Mean Square Error (RMSE), Heidke skill score (HSS), ranked probability skill score (RPSS), and a hit-miss matrix (Heidke, 1926; Epstein, 1969). HSS and RPSS are categorical performance metrics and can be interpreted as a percentage improvement over a reference forecast. A standard reference forecast for hydro-climate prediction is based on a climatological (equal odds) distribution of observed July–August cyanobacteria biomass. Here, the reference forecast is split into three categories of equal probability (33% each), representing *below normal* (0–2.91 mg/L), *near normal* (2.91–4.56 mg/L), and *above normal* (4.56+ mg/L) cyanobacteria conditions. For the forecast model developed here, if there is no predictive information, the model defaults to equal odds categorical prediction, as in the reference forecast. However, for most years, the distribution of expected conditions shifts and results in unequal likelihoods of each category. Thus the forecast developed here outperforms the reference forecast when it assigns a greater probability (more than 33%) to the category that is ultimately observed. The HSS is a deterministic categorical score:

$$HSS = \frac{\sum_i P(F_i, O_j) - \sum_i P(F_i)P(O_i)}{1 - \sum_i P(F_i)P(O_i)} \quad (3)$$

where the joint distribution of forecasts and observations is described by $P(F_i, O_j)$, and the marginal distributions of forecasts and observations are described by $P(F_i)$ and $P(O_j)$, respectively (Wilks 2011). HSS values range from $-\infty$ to 1, where 0 represents no skill and 1 represents a perfect forecast. RPSS is a probabilistic categorical score based on ensemble forecasts for each year (summertime) in the time series. RPSS uses the ranked probability score (RPS), a measure of the squared differences in the cumulative probability of a multi-categorical ensemble. The RPSS score is increasingly penalized as more forecast ensemble members are assigned to categories farther from the observed category. RPSS values range from $-\infty$ to 1, where 0 represents no skill, 1 represents a perfect forecast, and negative values indicate that forecasts are inferior to the reference forecast. RPSS values are generated for each year using Eq. 4;

the median value is reported.

$$RPSS = \frac{RPS - RPS_{reference}}{0 - RPS_{reference}} = 1 - \frac{RPS}{RPS_{reference}} \quad (4)$$

3. Results

3.1. Phase model performance

A unique set of cyanobacteria predictors are retained for the MEI positive and negative phase models (Table 1), validating the utility of separate models to describe this asymmetric relationship. Regions of May–June SST anomalies are identified following Zimmerman et al. (2016) for both the positive and negative phases (Fig. 3). In the negative phase (La Niña-like) model, significantly correlating regions of May–June SST anomalies are located in the equatorial Pacific Ocean. In the positive phase (El Niño-like) model, significantly correlating regions of May–June SST anomalies are located in the mid and northern Atlantic Ocean.

The final set of predictors for the negative (La Niña-like) phase includes June discharge, June phosphorus load, June total unfiltered phosphorus measured at the LTER buoy, the floating algae index for June, and May–June average SST anomalies in parts of the Pacific Ocean. These first four variables represent local-scale processes, and SSTs represent global scale processes, explaining cyanobacteria variability. The first three principal components are retained for the negative phase model and explain approximately 65%, 13%, and 10% of the variance, respectively.

The final set of predictors for the positive (El Niño-like) phase only includes May–June SST anomalies in the Atlantic Ocean and the floating algae index. The first principal component of the positive phase model explains approximately 92% of the variance in the data and is the only PC retained for the model. Variables commonly associated with cyanobacteria productivity (e.g. phosphorus, discharge, extreme precipitation events) are not statistically significant during the positive phase (Table 1).

3.2. Combined model performance

A hindcast assessment combining the positive and negative phase models results in Pearson and Spearman correlation coefficients of 0.90 and 0.83 respectively, an RMSE of 1.22, an HSS of 0.41, and a median RPSS of 0.72, indicating a clear improvement over climatology (Fig. 4). The model illustrates particular skill in predicting *below normal* and *above normal* conditions but performs poorly in the *near normal* category (Table 2). The model's ability to correctly predict *above normal* July–August cyanobacteria biomass (6 out of the 8 years) is particularly advantageous from a management perspective. Additionally, the two-phase model demonstrates substantial improvement over a traditional model that does not discriminate between ENSO phases (Fig. 5; the “one-phase” model results in Pearson and Spearman correlation coefficients of 0.81 and 0.70, respectively, an RMSE of 1.53, an HSS of 0.35, and a median RPSS of 0.56.)

3.3. Seasonal and sub-seasonal model comparison

A comparison between the full (June–Aug) and sub-seasonal (July–Aug) model outputs is warranted to understand agreement between models and potential gains from issuing an updated forecast. A probability density plot of the sub-seasonal hindcast appears to more accurately reflect observed conditions than the seasonal hindcast and illustrates the sub-seasonal model's increased accuracy in the tails, with less emphasis on the *near normal* category (Fig. 6). To assess the degree of difference between the predicted probability distributions and the observed distribution, two, two-sample Komolgorov-Smirnov test are performed. The Komolgorov-Smirnov test statistic (D) quantifies the

Table 1

Pearson and Spearman correlation coefficients between July-August average cyanobacteria biomass with ENSO phase indicated. Bold values indicate the set of significantly correlated predictors selected with the GCV for each phase model. * indicate significantly correlated variables.

Predictor	Months	ENSO Phase	Pearson	Spearman	Source
Discharge (USGS Station 05427718)	June	Positive	-0.09	-0.04	USGS
Phosphorus Load (USGS Station 05427718)	June		0.11	-0.13	USGS
Suspended Sediment Load (USGS Station 05427718)	June		0.22	-0.11	USGS
Soil Moisture (Grid: 43.313 -89.313)	June		0.15	0.25	NLDAS
Nitrate + Nitrite (Buoy)	June		-0.41	-0.21	LTER
Total Unfiltered Phosphorus (Buoy)	June		0.21	0.04	LTER
Sea Surface Temperature (PC1)	May-June		-0.87*	-0.96*	IRI Data Library
Sea Level Pressure (PC1)	May-June		-0.88*	-0.96*	IRI Data Library
Extreme Events (>25mm)	March-May		0.25	0.25	MRCC
Air Temperature	June		-0.11	-0.54	MRCC
Extreme Air Temperature Events	March-June		-0.002	-0.23	MRCC
Floating Algae Index	June		0.93*	0.89*	MODIS/Landsat
D. Pulicaria Biomass	June		-0.89*	-0.89*	LTER
Water Temperature	April-June		0.64	0.57	LTER
Pre-season Cyanobacteria Biomass	June		0.52	0.02	LTER
Discharge (USGS Station 05427718)	June	Negative	0.83*	0.63*	USGS
Phosphorus Load (USGS Station 05427718)	June		0.82*	0.57*	USGS
Suspended Sediment Load (USGS Station 05427718)	June		0.78*	0.55*	USGS
Soil Moisture (Grid: 43.313 -89.313)	June		0.43	0.43	NLDAS
Nitrate + Nitrite (Buoy)	June		0.63*	0.66*	LTER
Total Unfiltered Phosphorus (Buoy)	June		0.62*	0.61*	LTER
Sea Surface Temperature (PC1)	May-June		-0.74*	-0.57*	IRI Data Library
Sea Level Pressure (PC1)	May-June		-0.63	-0.41	IRI Data Library
Extreme Events (>25mm)	March-May		0.72*	0.75*	MRCC
Air Temperature	June		-0.05	-0.003	MRCC
Extreme Air Temperature Events	March-June		0.09	-0.02	MRCC
Floating Algae Index	June		0.71*	0.70*	MODIS/Landsat
D. Pulicaria Biomass	June		-0.15	-0.03	LTER
Water Temperature	April-June		0.58	0.39	LTER
Pre-season Cyanobacteria Biomass	June		0.18	0.60	LTER

distance between two empirical distribution functions. A smaller test statistic is found between the sub-seasonal and observed distributions ($D=0.22$) compared to the seasonal and observed distributions ($D=0.30$). Neither the predicted sub-seasonal or seasonal distribution was significantly different from the observed July-August biomass distribution at the 95% confidence level ($P>0.66$ and $P=0.23$, respectively). From a categorical perspective, normalized seasonal and sub-seasonal hindcasts correctly predict 56.5% and 60.8% of observed July-August biomass respectively (Table 3). Both hindcasts perform well in the *below normal* category and poorly in predicting *near normal* conditions. Most notably, the sub-seasonal prediction model for *above normal* cyanobacteria biomass is an improvement over the seasonal forecast model. Specifically, the sub-seasonal forecast correctly predicts *above normal* conditions for three years in which the seasonal forecast does not (Table 4). The sub-seasonal forecast incorrectly updated an *above normal* seasonal prediction in only one year (2011). Increased accuracy in prediction of *above normal* cyanobacteria conditions by the sub-seasonal forecast is encouraging, as these conditions present the greatest threat to public health.

4. Discussion

In addition to practical applications, prediction plays an important role in demonstrating ecological understanding (Houlahan et al., 2017). The development and assessment of the positive and negative phase forecasting models provides some insight into the relative importance of local and global scale variables on a seasonal timeframe.

At the local scale, the predictive power of several variables are noteworthy. The floating algae index, a remotely sensed indicator of pre-season lake productivity, is the only local-scale predictor significant in both phase models. Pre-season cyanobacteria biomass, however, was not a significant predictor of July-August biomass in either phase. This suggests that general algae productivity in the early summer may be more indicative of favorable conditions for July-August cyanobacteria

than early summer cyanobacteria biomass itself. Additionally, despite the established importance of temperature in cyanobacteria productivity, neither of the air temperature-based predictors are significantly correlated with July-August cyanobacteria biomass in either phase. Pre-season water temperatures resulted in higher correlation coefficients than air temperature predictors, but relationships were not strong enough to be included in either phase. Konopka and Brock (1978) purport that the relationship between lake temperature and cyanobacteria growth in Mendota is complicated by other concurrent environmental changes. Ultimately, the temperature-based predictors included here may be too simplistic to fully capture the relationships between air and water temperature and cyanobacteria growth.

At a global scale, regions of relevant sea surface temperatures identified by the NIPA process suggest differences in the influence of large-scale climate phenomena on local hydroclimatic processes in the Midwest during the positive and negative phases of ENSO. In the negative phase (La Niña-like) model, significantly correlating regions of May-June SST anomalies are located in the equatorial Pacific Ocean, a region commonly associated with ENSO (Fig. 3). A relationship has been previously established between springtime La Niña conditions and a strong Great Plains low level jet (GPLLJ), which acts as a conduit for moisture from the tropical Atlantic to the continental U.S. (Munoz and Enfield, 2011; Krishnamurthy et al., 2015). Increased springtime moisture during the MEI negative phase may explain why variables associated with precipitation (e.g. phosphorus load, discharge, extreme events, suspended sediments) are significantly correlated with July-August cyanobacteria biomass in the negative phase, but not in the positive phase (Table 1).

In the positive phase (El Niño-like) model, significantly correlating regions of May-June SST anomalies are located in the mid and northern Atlantic Ocean (Fig. 3). The GPLLJ draws moisture from the tropical Atlantic via the Caribbean low-level jet, however, Krishnamurthy et al. (2015) suggest that El Niño conditions are not typically associated with a strong GPLLJ in boreal spring (April-June). This may explain why

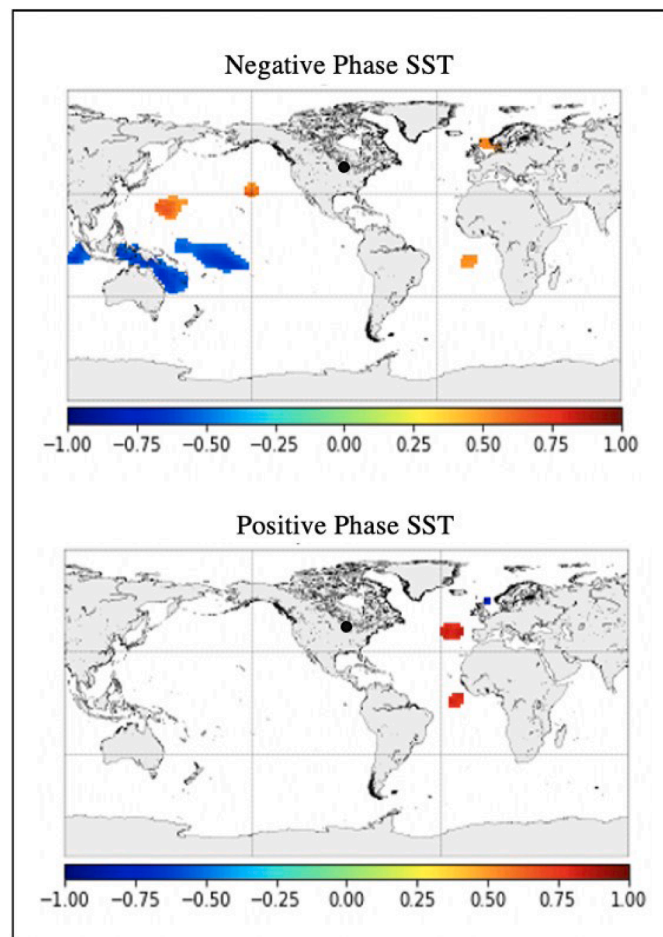


Fig. 3. Regions of statistically significant (95th-percentile) Pearson correlation coefficients between July-August cyanobacteria biomass and May-June SST anomalies for negative and positive ENSO phases. The black dot represents the study site. Colors represent the degree of correlation.

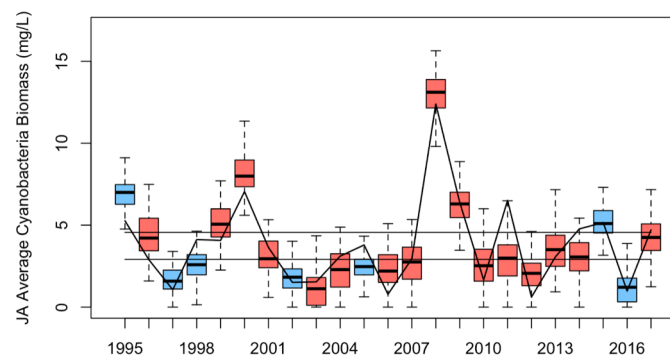


Fig. 4. July-August average cyanobacteria biomass predictions for positive and negative phases of ENSO (box plots) and observed data (solid black line.) Thresholds between *below*, *near*, and *above* normal categories are denoted by horizontal black lines.

regions of significantly correlating SST (and SLP) anomalies are focused in the Atlantic, and why they are absent from the equatorial Pacific Ocean. On average, total June precipitation is lower during the positive phase compared to the negative phase (Fig. 7). Furthermore, drought periods in the Yahara watershed have been shown to decrease July-August discharge, phosphorus loads, and total phosphorus within the lake (Lathrop and Carpenter, 2014). A weaker GPLJ in the positive

phase may explain lower June precipitation and the lack of significant correlations between precipitation-driven variables and cyanobacteria biomass in these years.

ENSO signals may also explain asymmetries in predictor relationships at the lake scale. *D. pulicaria* biomass is not selected by the model for the final suite of predictors but is significantly correlated with cyanobacteria biomass in the positive phase. Plankton community

Table 2

Cyanobacteria biomass forecast results: observed cyanobacteria biomass category vs. the forecasted category in a given year. Values represent the number of historical years that fall into each category based on a hindcast.

Forecast		Observed		
		Below Normal	Near Normal	Above Normal
Forecast	Below Normal	7	5	2
	Near Normal	1	1	0
	Above Normal	0	1	6

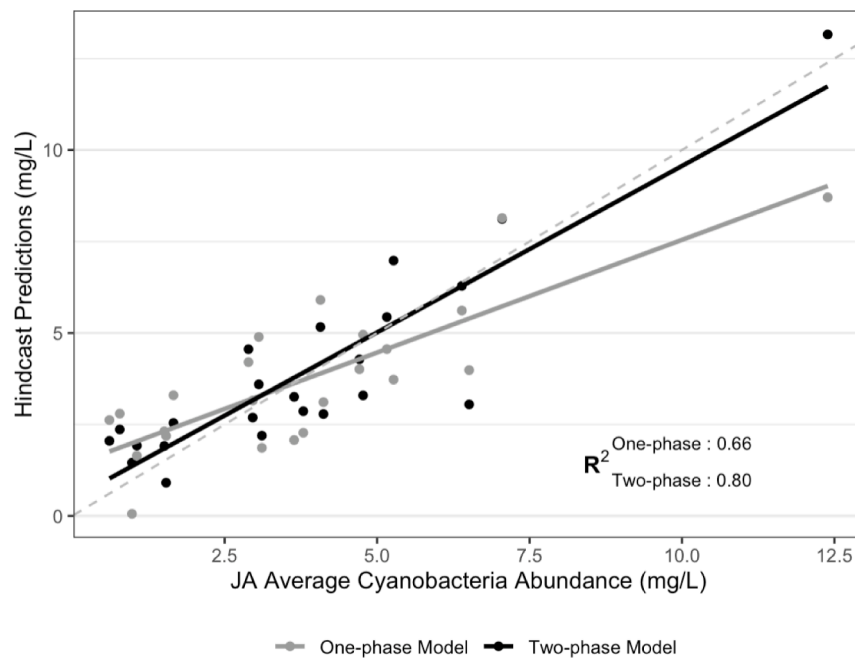


Fig. 5. July-August average cyanobacteria biomass observations compared to hindcast predictions using “one-phase” and “two-phase” (NIPA) models. A 1:1 line (perfect forecast) is represented by the dashed gray line.

dynamics in Lake Mendota are complex, however, there is some evidence that the effects of *D. pulicaria* grazing on phytoplankton are more pronounced when phosphorus concentrations are low (Vanni et al., 1990). It is possible then, that the influence of *D. pulicaria* grazing on July-August cyanobacteria biomass is more pronounced in the positive phase due to differences in phosphorus conveyance between the phases of ENSO.

In comparison to a model conditioned on all available years of cyanobacteria biomass data, the split phase model approach significantly improves predictions of July-August cyanobacteria biomass, most notably for *above normal* cyanobacteria conditions (Fig. 5). Additionally, the subseasonal forecast developed here shows an improvement in forecast skill for July-August cyanobacteria biomass over the full season (June-August) forecast developed by Soley et al. (2016) (Table 4, Fig. 6). This suggests that a sub-seasonal cyanobacteria forecast (released on July 1st) can provide lake and beach managers with a meaningful update to the full season forecast (released on June 1st), with greater accuracy regarding the peak months for cyanobacteria productivity in Lake Mendota.

Despite the sub-seasonal model's ability to improve forecasts of *above normal* cyanobacteria conditions overall, the model incorrectly updated a seasonal forecast of *above normal* cyanobacteria biomass to *near normal* in 2011 (Table 4). Increased cyanobacteria abundance in 2011 may have resulted from high concentrations of dissolved inorganic nitrogen throughout the summer. Beversdorf et al. (2013) suggest that an abnormally high ratio of dissolved inorganic nitrogen to dissolved reactive phosphorus in 2011 allowed the early summer cyanobacteria

species *Aphanizomenon* to persist into July and August, coexisting with *Mycrocystis*, a species typical of later summer months. It is possible that cyanobacteria biomass in 2011 was driven primarily by nitrogen availability, a predictor not selected for the negative phase model by the GCV. Neglecting to distinguish between cyanobacteria species may limit the model's ability to capture the influence of community dynamics on overall cyanobacteria abundance throughout the summer.

While linear models are unable to entirely capture non-linear drivers of atmospheric and limnological processes, the NIPA approach highlights the diverse response of local and global predictor variables important to cyanobacteria productivity given the mean state of the atmospheric-oceanic system. Differences in the predictive power of phosphorus load and related variables (e.g. discharge, extreme events, suspended sediments) between MEI phases are particularly notable considering the large body of work establishing phosphorus load as a major driver of cyanobacteria productivity in Lake Mendota. The number of years in each phase is relatively small from a statistical perspective, however, compared to most inland lakes, the record for Lake Mendota is considered long. Nonetheless, continued collection of water quality data is warranted for the refinement of these models.

5. Conclusions

In this paper, skillful sub-seasonal cyanobacteria biomass prediction models are developed and compared with full-season prediction models to understand potential prediction gains and inform lake and beach management. The inter-annual variability of biomass results from a

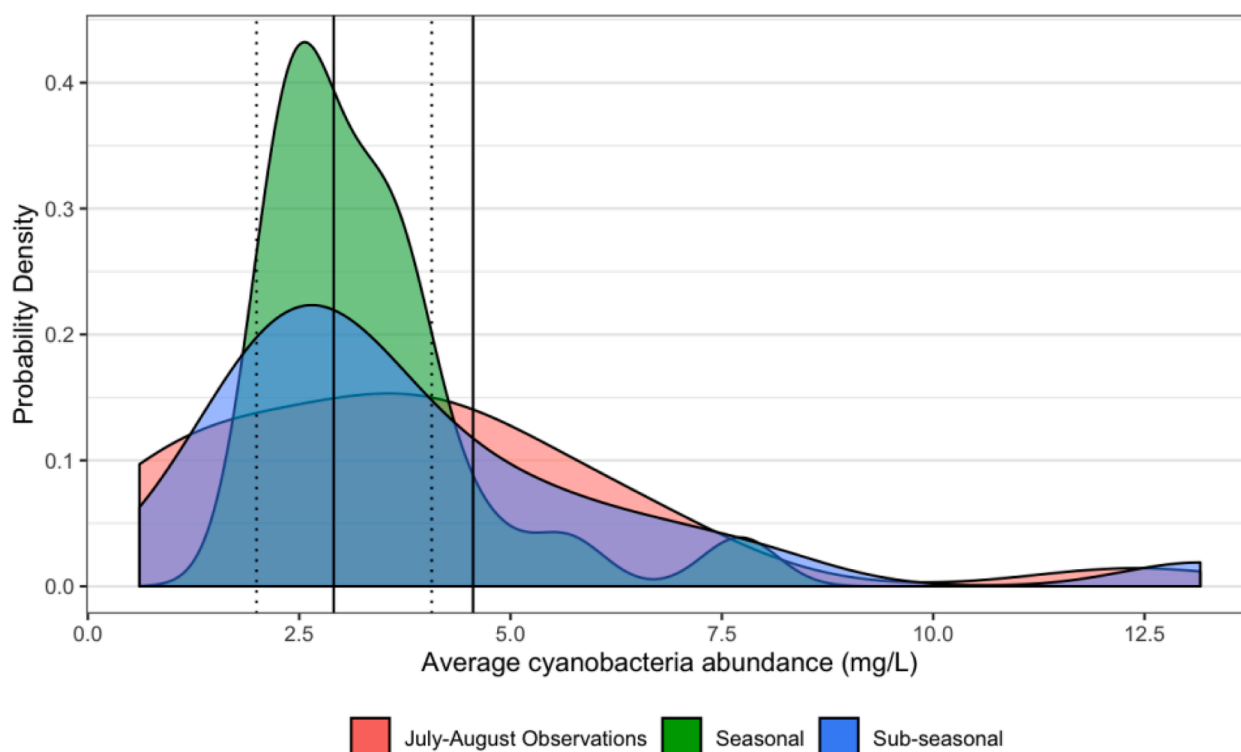


Fig. 6. PDFs of average cyanobacteria biomass observations (July-August), seasonal prediction (June-August), and sub-seasonal prediction (July-August across 1995–2017), represented by red, green, and blue areas, respectively. Additional colors represent areas in which the PDFs overlap. Dashed vertical lines indicate thresholds between *below*, *near*, and *above normal* seasonal (June-August) categories. Solid lines indicate sub-seasonal (July-August) category thresholds.

Table 3

Observations and number of correct normalized categorical predictions of July-August cyanobacteria biomass from 1995 to 2017.

Category	Years observed	Correct Seasonal Forecasts (%)	Correct Sub-seasonal Forecasts (%)
Above Normal	8	3 (37.5)	5 (62.5)
Near Normal	7	2 (28.6)	2 (28.6)
Below Normal	8	8 (100)	7 (87.5)
All	23	13 (56.5)	14 (60.8)

Table 4

Years in which the sub-seasonal forecast corrected an incorrect seasonal forecast (Corrections) and years in which the sub-seasonal forecast miscorrected an accurate seasonal forecast (Miscorrections). “Corrections” always imply an incorrect seasonal forecast. “Miscorrections” imply a correct seasonal forecast. Cases in which both forecasts are correct or incorrect are not represented.

Category	Corrections	Miscorrections
Above Normal	3	1
Near Normal	2	2
Below Normal	0	1

complex array of physical, chemical, and biological variables, many of which are significantly impacted by local climate, yet modulated broadly by large-scale climate phenomena through atmospheric teleconnections such as ENSO. In this paper, spring and early summer variables are evaluated to determine their ability to represent within-season drivers of July-August cyanobacteria biomass.

In comparison to a traditional model conditioned on all years in the historical record, a two-phase approach is adopted – categorizing years as falling into either a positive or negative phase according to the pre-season MEI value. This modeling approach significantly improves predictions of July-August cyanobacteria biomass – particularly for *above normal* July-August conditions – and highlights the relative importance of unique local and global cyanobacteria biomass drivers in each phase.

Notably, variables closely related to spring and summer phosphorus load are included in the negative phase model however are not significantly correlated with cyanobacteria biomass in the positive phase. This distinct behavior difference may be mediated by atmospheric teleconnections between ENSO and the Great Plains Low-Level Jet, which acts as a conduit for moisture transport from the mid-Atlantic to the Midwest. While inferences in how precipitation and thus variability in lake processes is modulated by ENSO specifically and large-scale climate generally are provided here, additional investigation is still warranted (Justić et al., 2005; Morse et al., 2014). Additional lines of inquiry could include development of coupled seasonal and sub-seasonal forecast systems for other water quality indicators, use of remote sensing methods to enhance observational records and predictability, and further integration of forecasts with lake and beach management alternatives.

Harmful algae

Author declaration

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work

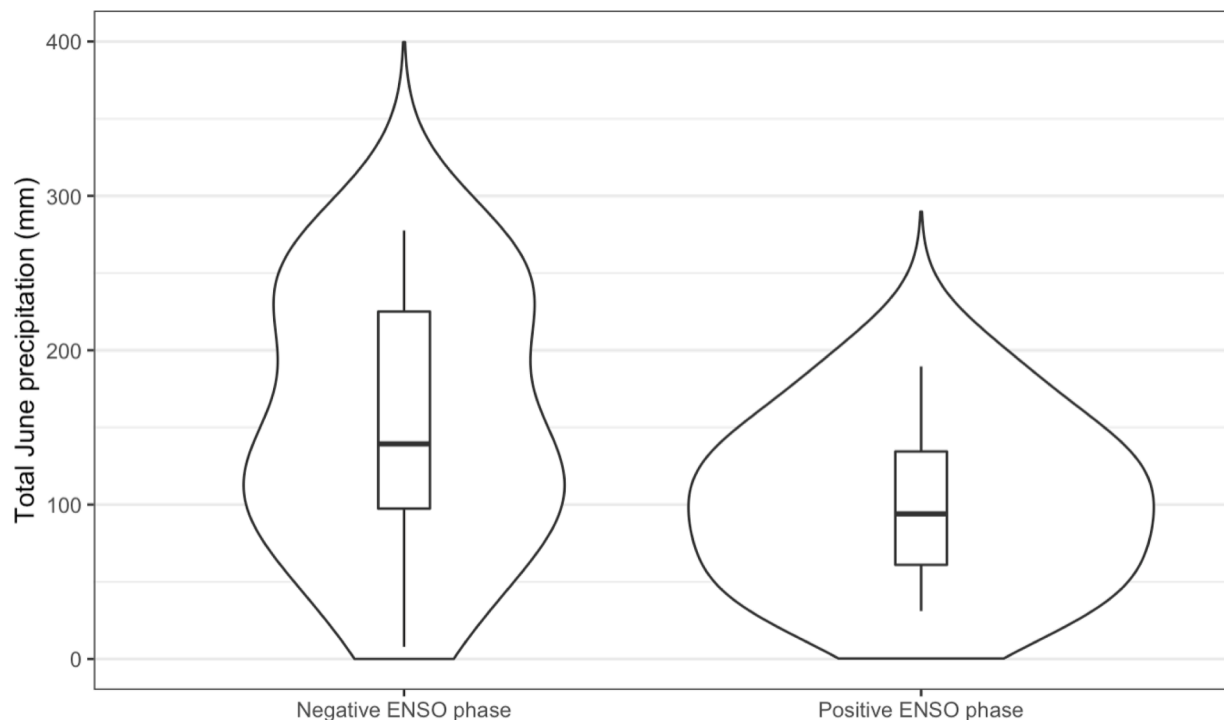


Fig. 7. Violin plot of total June precipitation for positive and negative ENSO phases (1995–2017) Mean precipitation is 153.8 and 100.7 millimeters for each phase, respectively. Precipitation data recorded at the Dane County Regional Airport was obtained from the Midwest Regional Climate Center for March–May (Wuertz et al., 2020).

was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, without the written consent of the copyright-holder.

By attaching this Declaration to the submission, the corresponding author certifies that:

- The manuscript represents original and valid work and that neither this manuscript nor one with substantially similar content under the same authorship has been published or is being considered for publication elsewhere.
- Every author has agreed to allow the corresponding author to serve as the primary correspondent with the editorial office, and to review the edited typescript and proof.
- Each author has given final approval of the submitted manuscript and order of authors. Any subsequent change to authorship will be approved by all authors.
- Each author has participated sufficiently in the work to take public responsibility for all the content.

Declaration of Competing Interest

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

Acknowledgments

Funding to support this work has been partially provided by an NSF CAREER program (Award Number: 1845783). Additional support was provided by the NSF Northern Temperate Lakes Long-term Ecological Research Program (DEB-1440297). Datasets for this research are available in these in-text data citation references: Allan and Ansel (2006), Menne et al. (2012), Mocko et al. (2013), Robertson (2016), Magnuson et al. (2019), Magnuson et al. (2020a), Magnuson et al. (2020b), Magnuson et al. (2020c), ORNL DAAC (2020), Smith et al. (2008), U.S.

Geological Survey (2020a), U.S. Geological Survey (2020b), and Wuertz et al. (2020). The model code used in this study is available at <https://github.com/mrwbeal/MendotaCyanobacteriaForecast>.

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. Manag.* 145 (5), 04019006.
- Allan, R., Ansell, T., 2006. A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J. Climate* 19, 5816–5842. <https://doi.org/10.1175/JCLI3937.1>.
- Anderson, W.L., Robertson, D.M., Magnuson, J.J., 1996. Evidence of recent warming and El Niño-related variations in ice breakup of Wisconsin lakes. *Limnol. Oceanogr.* 41 (5), 815–821.
- Anneville, O., Domaizon, I., Kerimoglu, O., Rimet, F., Jacquet, S., 2015. Blue-green algae in a “Greenhouse Century”? New insights from field data on climate change impacts on cyanobacteria abundance. *Ecosystems* 18 (3), 441–458.
- Aoki, I., 1989. Holological study of lakes from an entropy viewpoint-lake Mendota. *Ecol. Model.* 45 (2), 81–93.
- Atech, G., 2000. Cost of Algal Blooms. Land and Water Resources Research and Development Corporation, Canberra, Australia.
- Baker, S.A., Wood, A.W., Rajagopalan, B., 2019. Developing subseasonal to seasonal climate forecast products for hydrology and water management. *JAWRA J. Am. Water Resour. Assoc.* 55 (4), 1024–1037.
- Barnston, A.G., 1994. Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Clim.* 7 (10), 1513–1564.
- Beversdorf, L.J., Miller, T.R., McMahon, K.D., 2013. The role of nitrogen fixation in cyanobacterial bloom toxicity in a temperate, eutrophic lake. *PLoS One* 8 (2), e56103.
- Block, P.J., Filho, Souza, F., A., Sun, L., Kwon, H.H., 2009. A streamflow forecasting framework using multiple climate and hydrological models 1. *JAWRA J. Am. Water Resour. Assoc.* 45 (4), 828–843.
- Brezonik, P.L., Lee, G.F., 1968. Denitrification as a nitrogen sink in Lake Mendota, Wisconsin. *Environ. Sci. Technol.* 2 (2), 120–125.
- Briand, J.F., Lebourlangier, C., Humbert, J.F., Bernard, C., Dufour, P., 2004. *Cylindrospermopsis raciborskii* (cyanobacteria) invasion at mid-latitudes: selection, wide physiological tolerance, or global warming? 1. *J. Phycol.* 40 (2), 231–238.
- Brock, T.D., 1985. A Eutrophic Lake: Lake Mendota, Wisconsin (Vol. 55). Springer Science & Business Media, Berlin, Germany.
- Carmichael, W.W., 1994. The toxins of cyanobacteria. *Sci. Am.* 270 (1), 78–86.
- 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.

- 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 (1), 71–79.
- Carpenter, S.R., Booth, E.G., Kucharik, C.J., 2018. Extreme precipitation and phosphorus loads from two agricultural watersheds. *Limnol. Oceanogr.* 63 (3), 1221–1233.
- Centers for Disease Control and Prevention (CDC), 2020. National Outbreak Reporting System Dashboard. U.S. Department of Health and Human Services, CDC, Atlanta, Georgia. Retrieved August 20 from www.cdc.gov/norsdashboard.
- Chiew, F.H.S., Zhou, S.L., McMahon, T.A., 2003. Use of seasonal streamflow forecasts in water resources management. *J. Hydrol.* 270 (1–2), 135–144.
- Craven, P., Wahba, G., 1978. Smoothing noisy data with spline functions. *Numer. Math.* 31 (4), 377–403.
- Delorit, J., Gonzalez, E., Block, P., 2017. Evaluation of model-based seasonal streamflow and water allocation forecasts for the Elqui Valley, Chile. *Hydrol. Earth Syst. Sci.* 21, 4711–4725. <https://doi.org/10.5194/hess-21-4711-2017>.
- Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Thornbrugh, D.J., 2009. Eutrophication of US freshwaters: analysis of potential economic damages. *Environ. Sci. Technol.* 43 (1), 12–19. <https://doi.org/10.1021/es801217q>.
- Downing, J.A., Watson, S.B., McCauley, E., 2001. Predicting cyanobacteria dominance in lakes. *Can. J. Fish. Aquat. Sci.* 58 (10), 1905–1908.
- Elliott, J.A., 2012. Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. *Water Res.* 46 (5), 1364–1371.
- Epp, G.T., 1996. Grazing on filamentous cyanobacteria by *Daphnia pulicaria*. *Limnol. Oceanogr.* 41 (3), 560–567.
- Epstein, E.S., 1969. A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol.* 8 (6), 985–987.
- Fu, W., Steinschneider, S., 2019. A diagnostic-predictive assessment of winter precipitation over the Laurentian great lakes: effects of ENSO and other teleconnections. *J. Hydrometeorol.* 20 (1), 117–137.
- Gallina, N., Anneville, O., Beniston, M., 2011. Impacts of extreme air temperatures on cyanobacteria in five deep peri-Alpine lakes. *Limnol. Oceanogr.* 70 (2), 186–196.
- Garnache, C., Swinton, S.M., Herriges, J.A., Lupi, F., Stevenson, R.J., 2016. Solving the phosphorus pollution puzzle: synthesis and directions for future research. *Am. J. Agric. Econ.* 98 (5), 1334–1359.
- Genskow, K., Betz, C., 2012. Farm practices in the Lake Mendota Watershed: a comparative analysis of 1996 and 2011.
- Giannini, A., Kushnir, Y., Cane, M.A., 2000. Interannual variability of Caribbean rainfall, ENSO, and the Atlantic Ocean. *J. Clim.* 13 (2), 297–311.
- Gobler, C.J., Burkholder, J.M., Davis, T.W., Harke, M.J., Johengen, T., Stow, C.A., Van de Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. *Harmful Algae* 54, 87–97.
- Giuliani, M., Zaniolo, M., Castelletti, A., Davoli, G., Block, P., 2019a. Detecting the state of the climate system via artificial intelligence to improve seasonal forecasts and inform reservoir operations. *Water Resour. Res.* 55 (11), 9133–9147. <https://doi.org/10.1029/2019WR025035>.
- Giuliani, M., Zaniolo, M., Castelletti, A., Block, P., Zimmerman B., Carlino, A., Amaranto, A., 2019b. Climate state intelligence. 2019b. <https://github.com/mxgiuliani00/CSI>.
- Guo, L., 2007. Doing battle with the green monster of Taihu Lake. *Science* 317 (5842), 1166.
- Håkanson, L., Bryhn, A.C., Hytteborn, J.K., 2007. On the issue of limiting nutrient and predictions of cyanobacteria in aquatic systems. *Sci. Total Environ.* 379 (1), 89–108.
- Hart, M.R., Quin, B.F., Nguyen, M.L., 2004. Phosphorus runoff from agricultural land and direct fertilizer effects: a review. *J. Environ. Qual.* 33 (6), 1954–1972.
- Heidke, P., 1926. Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler* 8 (4), 301–349.
- Helsel, D.R., Hirsch, R.M., 1992. Statistical Methods in Water Resources. Elsevier Science Publishers B.V.
- Houlahan, J.E., McKinney, S.T., Anderson, T.M., McGill, B.J., 2017. The priority of prediction in ecological understanding. *Oikos* 126 (1), 1–7.
- Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* 113 (10), 2118–2129.
- Huber, V., Wagner, C., Gerten, D., Adrian, R., 2012. To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. *Oecologia* 169 (1), 245–256.
- Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M., Visser, P.M., 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* 16 (8), 471–483.
- Justić, D., Rabalais, N.N., Turner, R.E., 2005. Coupling between climate variability and coastal eutrophication: evidence and outlook for the northern Gulf of Mexico. *J. Sea Res.* 54 (1), 25–35.
- Kasprzak, P.H., Lathrop, R.C., 1997. Influence of two *Daphnia* species on summer phytoplankton assemblages from eutrophic lakes. *J. Plankton Res.* 19 (8), 1025–1044.
- Kahya, E., Dracup, J.A., 1993. US streamflow patterns in relation to the El Niño/Southern Oscillation. *Water Resour. Res.* 29 (8), 2491–2503.
- Konopka, A., Brock, T.D., 1978. Effect of temperature on blue-green algae (cyanobacteria) in Lake Mendota. *Appl. Environ. Microbiol.* 36 (4), 572–576.
- Krishnamurthy, L., Vecchi, G.A., Msadek, R., Wittenberg, A., Delworth, T.L., Zeng, F., 2015. The seasonality of the Great Plains low-level jet and ENSO relationship. *J. Clim.* 28 (11), 4525–4544.
- Kronvang, B., Jeppesen, E., Conley, D.J., Søndergaard, M., Larsen, S.E., Ovesen, N.B., Carstensen, J., 2005. Nutrient pressures and ecological responses to nutrient loading reductions in Danish streams, lakes and coastal waters. *J. Hydrol.* 304 (1–4), 274–288.
- Lathrop, R.C., Carpenter, S.R., 1992. Phytoplankton and their relationship to nutrients. *Food Web Management*. Springer, New York, NY, pp. 97–126.
- Lathrop, R.C., Carpenter, S.R., Rudstam, L.G., 1996. Water clarity in Lake Mendota since 1900: responses to differing levels of nutrients and herbivory. *Can. J. Fish. Aquat. Sci.* 53 (10), 2250–2261.
- 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 (5), 1169–1178.
- Lathrop, R.C., Carpenter, S.R., Robertson, D.M., 1999. Summer water clarity responses to phosphorus, *Daphnia* grazing, and internal mixing in Lake Mendota. *Limnol. Oceanogr.* 44 (1), 137–146.
- Lathrop, R.C., 2007. Perspectives on the eutrophication of the Yahara lakes. *Lake Reservoir Manag.* 23 (4), 345–365.
- Lathrop, R.C., Carpenter, S.R., 2014. Water quality implications from three decades of phosphorus loads and trophic dynamics in the Yahara chain of lakes. *Inland Waters* 4 (1), 1–14.
- Lee, D., Ward, P.J., Block, P., 2018. Identification of symmetric and asymmetric responses in seasonal streamflow globally to ENSO phase. *Environ. Res. Lett.* 13 (4), 044031.
- Magnuson J., Carpenter S., Stanley E., 2019. North Temperate Lakes LTER: Zooplankton – Madison Lakes Area 1997 - current ver 31. *Environ. Data Initiat.* 10.6073/pasta/8b265c0300252c87805f26f41e174aa4. Accessed 2020-06-09.
- Magnuson, J., Carpenter, S., Stanley, E., 2020a. North temperate lakes LTER: phytoplankton - Madison Lakes Area 1995 - current ver 28. *Environ. Data Initiat.* <https://doi.org/10.6073/pasta/13ea8f578654493155a660ab886f695e>. Accessed 2020-06-09.
- Magnuson, J., Carpenter, S., Stanley, E., 2020b. North temperate lakes LTER: chemical limnology of primary study lakes: nutrients, pH and carbon 1981 - current ver 52. *Environ. Data Initiat.* <https://doi.org/10.6073/pasta/8359d27bbd91028f222d923a7936077d>. Accessed 2020-06-09.
- Magnuson, J., Carpenter, S., Stanley, E., 2020c. North temperate lakes LTER: high frequency water temperature data - Lake Mendota Buoy 2006 - current ver 29. *Environ. Data Initiat.* <https://doi.org/10.6073/pasta/8ceff296ad68fa8da6787076e0a5d992>. Accessed 2021-04-14.
- Markowski, G.R., North, G.R., 2003. Climatic influence of sea surface temperature: evidence of substantial precipitation correlation and predictability. *J. Hydrometeorol.* 4 (5), 856–877.
- Menne, M.J., Durre, I., Korzeniewski, B., McNeal, S., Thomas, K., Yin, X., Anthony, S., Ray, R., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. Global Historical Climatology Network-Daily (GHCN-Daily), Version 3. NOAA National Climatic Data Center. <https://doi.org/10.7289/V5D21VHZ>. Accessed: 2020-09-16. Retrieved from: <https://mrcc.illinois.edu/CLIMATE/Station/Daily/StnDyBTD.jsp>.
- Michalak, A.M., 2016. Study role of climate change in extreme threats to water quality. *Nature* 535 (7612), 349–350.
- Mocko, D., NASA/GSFC/HSL, 2013. NLDAS Noah Land Surface Model L4 Monthly Climatology 0.125 x 0.125 degree V002. Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA. <https://doi.org/10.5067/USBAFY8R76IK>. Accessed: 2020-09-16.
- Morse, N.B., Wollheim, W.M., 2014. Climate variability masks the impacts of land use change on nutrient export in a suburbanizing watershed. *Biogeochemistry* 121 (1), 45–59.
- ... Motew, M., Chen, X., Booth, E.G., Carpenter, S.R., Pinkas, P., Zipper, S.C., Kucharik, C. J., 2017. The influence of legacy P on lake water quality in a Midwestern agricultural watershed. *Ecosystems* 20 (8), 1468–1482.
- Munoz, E., Enfield, D., 2011. The boreal spring variability of the Intra-Americas low-level jet and its relation with precipitation and tornadoes in the eastern United States. *Clim. Dyn.* 36 (1–2), 247–259.
- ORNL DAAC, 2020. MODIS and VIIRS Land Products Global Subsetting and Visualization Tool. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed September 16, 2020. Subset obtained for MCD43A4.006 MODIS Nadir BRDF-Adjusted Reflectance Daily 500m product at 43.0989, -89.4055, time period: 6-1-2020 to 6-26-2020, and subset size: 0.5 x 0.5km. 10.3334/ORNLDAAC/1379. Accessed 2020-10-02.
- Oyama, Y., Fukushima, T., Matsushita, B., Matsuzaki, H., Kamiya, K., Kobinata, H., 2015. Monitoring levels of cyanobacterial blooms using the visual cyanobacteria index (VCI) and floating algae index (FAI). *Int. J. Appl. Earth Obs. Geoinf.* 38, 335–348.
- Paerl, H.W., 1988. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. *Limnol. Oceanogr.* 33 (4–2), 823–843.
- Paerl, H.W., Fulton, R.S., Moisan, P.H., Dyble, J., 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *Sci. World J.* 1, 76–113.
- Paerl, H.W., Huisman, J., 2008. Blooms like it hot. *Science* 320 (5872), 57–58.
- Patel, J., Parshina-kottas, Y., 2017, October 03. Miles of algae covering Lake Erie. Retrieved August 20, 2020, from <https://www.nytimes.com/interactive/2017/10/03/science/earth/lake-erie.html>.
- Rabalais, N.N., Diaz, R.J., Levin, L.A., Turner, R.E., Gilbert, D., Zhang, J., 2010. Dynamics and distribution of natural and human-caused hypoxia. *Biogeosciences* 7 (2), 585.
- Rastetter, E.B., 2017. Modeling for understanding v. modeling for numbers. *Ecosystems* 20 (2), 215–221.
- Recknagel, F., Orr, P.T., Bartkow, M., Swanepoel, A., Cao, H., 2017. Early warning of limit-exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. *Harmful Algae* 69, 18–27.
- Regonda, S.K., Rajagopalan, B., Clark, M., Zagana, E., 2006. A multimodel ensemble forecast framework: application to spring seasonal flows in the Gunnison River Basin. *Water Resour. Res.* 42 (9).
- Reynolds, C.S., Oliver, R.L., Walsby, A.E., 1987. Cyanobacterial dominance: the role of buoyancy regulation in dynamic lake environments. *N.Z. J. Mar. Freshwater Res.* 21 (3), 379–390.

- 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 (3), 391–399.
- Robertson, D., 2016. Lake Mendota water temperature secchi depth snow depth ice thickness and meteorological conditions 1894 - 2007 ver 1. *Environ. Data Initiat.* 10.6073/pasta/f20f9a644bd12e4b80cb288f1812c935.
- Robertson, D.M., Wynne, R.H., Chang, W.Y., 2000. Influence of El Niño on lake and river ice cover in the Northern Hemisphere from 1900 to 1995. *Int. Vereinigung Theor. Angew. Limnol.* 27 (5), 2784–2788.
- Ropelewski, C.F., Halpert, M.S., 1986. North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Month. Weather Rev.* 114 (12), 2352–2362.
- Ropelewski, C.F., Halpert, M.S., 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Month. Weather Rev.* 115 (8), 1606–1626.
- Sarnelle, O., 2007. Initial conditions mediate the interaction between Daphnia and bloom-forming cyanobacteria. *Limnol. Oceanogr.* 52 (5), 2120–2127.
- Shentsis, I., Ben-Zvi, A., 1999. Within-season updating of seasonal forecast of Lake Kinneret inflow. *J. Hydrol. Eng.* 4 (4), 381–385.
- Schindler, D.W., 1977. Evolution of phosphorus limitation in lakes. *Science* 195 (4275), 260–262.
- Shabbar, A., Skinner, W., 2004. Summer drought patterns in Canada and the relationship to global sea surface temperatures. *J. Clim.* 17 (14), 2866–2880.
- Sinha, E., Michalak, A.M., Balaji, V., 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science* 357 (6349), 405–408.
- Smith, V.H., 1985. Predictive models for the biomass of blue-green algae in lakes 1. *JAWRA J. Am. Water Resour. Assoc.* 21 (3), 433–439.
- Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. *Environ. Sci. Pollut. Res.* 10 (2), 126–139.
- Smith, T.M., Reynolds, R.W., Peterson, T.C., Lawrimore, J., 2008. Improvements NOAA's historical merged land-ocean temp analysis (1880–2006). *J. Clim.* 21, 2283–2296.
- Stern, M., Kornfield, M., 2016, June 10. Why Florida's toxic algae crisis is worse than people realize. Retrieved August 20, 2020, from <https://www.tampabay.com/news/environment/2020/06/08/why-floridas-toxic-algae-crisis-is-worse-than-peoples-realize/>.
- 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 (6), 1174–1183.
- Soley, C., 2016. Cyanobacteria Abundance Modeling: Development and Assessment of Season-Ahead Forecasts to Improve Beach Management on Lake Mendota (MS Thesis). University of Wisconsin-Madison, Madison, WI. Retrieved from MINDS@UW. <https://minds.wisconsin.edu/handle/1793/75363>.
- Søndergaard, M., Jensen, J.P., Jeppesen, E., 2003. Role of sediment and internal loading of phosphorus in shallow lakes. *Hydrobiologia* 506 (1–3), 135–145.
- Stow, C.A., Carpenter, S.R., Laphrop, R.C., 1997. A Bayesian observation error model to predict cyanobacterial biovolume from spring total phosphorus in Lake Mendota, Wisconsin. *Can. J. Fish. Aquat. Sci.* 54 (2), 464–473.
- ... Taranu, Z.E., Gregory-Eaves, I., Leavitt, P.R., Bunting, L., Buchaca, T., Catalan, J., Moorhouse, H., 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. *Ecol. Lett.* 18 (4), 375–384.
- Towler, E., Rajagopalan, B., Summers, R.S., Yates, D., 2010. An approach for probabilistic forecasting of seasonal turbidity threshold exceedance. *Water Resour. Res.* 46 (6).
- University of Maryland Center for Environmental Science., 2015, May 11. Harmful algal blooms in the Chesapeake Bay are becoming more frequent. *ScienceDaily*. Retrieved August 20, 2020 from www.sciencedaily.com/releases/2015/05/150511125219.htm.
- U.S. Geological Survey., 2020a. National Water Information System data available on the World Wide Web (USGS Water Data for the Nation). Accessed: 2020-09-16. Retrieved from https://waterdata.usgs.gov/usa/nwis/uv?site_no=05427718.
- U.S. Geological Survey., 2020b. Landsat-5 Surface Reflectance Tier 1. Accessed: 2020-09-16. Retrieved from <https://code.earthengine.google.com>.
- Vanni, M.J., Luecke, C., Kitchell, J.F., Magnuson, J.J., 1990. Effects of planktivorous fish mass mortality on the plankton community of Lake Mendota, Wisconsin: implications for biomanipulation. *Biomaniipulation Tool for Water Management*. Springer, New York, NY, pp. 329–336.
- Vitart, F., Robertson, A.W., Anderson, D.L., 2012. Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *Bull. World Meteorol. Organ.* 61 (2), 23.
- Vitart, F., 2014. Evolution of ECMWF sub-seasonal forecast skill scores. *Q. J. R. Meteorol. Soc.* 140 (683), 1889–1899.
- Vitart, F., Robertson, A.W., 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Clim. Atmos. Sci.* 1 (1), 1–7.
- Vollenweider, R.A., 1970. *Scientific Fundamentals of the Eutrophication of Lakes and Flowing Waters, With Particular Reference to Nitrogen and Phosphorus as Factors in Eutrophication*. OECD, Paris, France.
- Walsby, A.E., Hayes, P.K., Boje, R., Stal, L.J., 1997. The selective advantage of buoyancy provided by gas vesicles for planktonic cyanobacteria in the Baltic Sea. *New Phytol.* 136 (3), 407–417.
- Wang, Y., Zheng, T., Zhao, Y., Jiang, J., Wang, Y., Guo, L., Wang, P., 2013. Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. *Environ. Sci. Pollut. Res.* 20 (12), 8909–8923.
- Wilks, D.S., 2011. *Statistical Methods in the Atmospheric Sciences* (Vol. 100). Academic press, Cambridge, MA.
- Wuertz, D., Lawrimore, J., Korzeniewski, B., 2020. Cooperative Observer Program (COOP) Hourly Precipitation Data (HPD), Version 2.0. NWS COOP Number: 474961. NOAA National Centers for Environmental Information. <https://doi.org/10.25921/p7j8-2170>. Accessed: 2020-10-02. Retrieved from: <https://mrcc.illinois.edu/CLIMA/TE/Station/Daily/StnDyBTD.jsp>.
- 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.
- Xu, H., Paerl, H.W., Qin, B., Zhu, G., Gao, G., 2010. Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnol. Oceanogr.* 55 (1), 420–432.
- Zimmerman, B.G., Vimont, D.J., Block, P.J., 2016. Utilizing the state of ENSO as a means for season-ahead predictor selection. *Water Resour. Res.* 52 (5), 3761–3774.
- Zhang, M., Duan, H., Shi, X., Yu, Y., Kong, F., 2012. Contributions of meteorology to the phenology of cyanobacterial blooms: implications for future climate change. *Water Res.* 46 (2), 442–452.
- Zhang, X., Recknagel, F., Chen, Q., Cao, H., Li, R., 2015. Spatially-explicit modelling and forecasting of cyanobacteria growth in Lake Taihu by evolutionary computation. *Ecol. Model.* 306, 216–225.
- Zhang, Y., Moges, S., Block, P., 2016. Optimal cluster analysis for objective regionalization of seasonal precipitation in regions of high spatial-temporal variability: Application to western Ethiopia. *J. Clim.* 29 (10), 3697–3717. <https://doi.org/10.1175/JCLI-D-15-0582.1>.