Recent Changes in Cyanobacteria Algal Bloom Magnitude in Large Lakes across the Contiguous United States

Sachidananda Mishra[1, 2], Richard P. Stumpf[2], Blake A. Schaeffer[3], P. Jeremy Werdell[4]*

[1]Consolidated Safety Services Inc., Fairfax, VA 22030, USA

[2]National Oceanic and Atmospheric Administration, National Centers for Coastal Ocean Science, Silver Spring, MD 20910, USA

[3]U.S. Environmental Protection Agency, Office of Research and Development, Durham, NC 27709, USA

[4]Ocean Ecology Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

Abstract

Cyanobacterial blooms in inland lakes produce large quantities of biomass that impact drinking water systems, recreation, and tourism and may produce toxins that can adversely affect public health. This study analyzed nine years of satellite-derived bloom records and compared how the bloom magnitude has changed from 2008-2011 to 2016-2020 in 1881 of the largest lakes across the contiguous United States (CONUS). We determined bloom magnitude each year as the spatio-temporal mean cyanobacteria biomass from May to October and in concentrations of chlorophyll-a. We found that bloom magnitude decreased in 465 (25%) lakes in the 2016-2020 period. Conversely, there was an increase in bloom magnitude in only 81 lakes (4%). Bloom magnitude either didn't change, or the observed change was in the uncertainty range in the majority of the lakes (n=1335, 71%). Above-normal wetness and normal or below-normal maximum temperature over the warm season may have caused the decrease in bloom magnitude in the eastern part of the CONUS in recent years. On the other hand, a hotter and dryer warm season in the western CONUS may have created an environment for increased algal biomass. While more lakes saw a decrease in bloom magnitude, the pattern was not monotonic over the CONUS. The variations in temporal changes in bloom magnitude within and across climatic regions depend on the interactions between land use land cover (LULC) and physical factors such as temperature and precipitation. Despite expectations suggested by recent global studies, bloom magnitude has not increased in larger US lakes over this time period.

Keywords: CyanoHABs, lacustrine algal blooms, satellite, remote sensing

1. Introduction

Algal blooms are an emerging environmental issue adversely affecting and disrupting aquatic ecosystems globally (Brooks et al., 2016; Hou et al., 2022). Several non-toxic algal species can have high biomass, producing discoloration, hypoxia, and a foul odor that can adversely impact recreational activity, the economy, and ecosystems (Hallegraeff et al., 2021; Kudela et al., 2015). In addition, several species of cyanobacteria can produce cyano-toxins such as microcystins, anatoxins, cylindrospermopsin, and saxitoxins, all of which pose risks to human and animal health (Loftin et al., 2016). Intravenous exposure to microcystin caused an outbreak of acute liver failure and 76 deaths at a dialysis center in Caruaru, Brazil, in 1996 (Carmichael et al., 2001). Although there is no comprehensive estimate of global economic loss due to harmful algal blooms (HABs), one study conservatively estimates the financial loss at several billion dollars (Kudela et al., 2015). A case study on the socio-economic impact, predominantly healthcare costs, of a single cyanobacteria harmful algal bloom (cyanoHAB) event in Utah Lake, USA, was valued at approximately \$370,000 (2017 U.S. dollars) in 2017 (Stroming et al., 2020). Additionally, the frequent occurrence of cyanoHABs in inland lakes can affect the housing market. Zhang et al. (Zhang et al., 2022) reported that more frequent cyanoHABs in lakes or nearby water bodies decreased property values in four climate regions (Upper Midwest, South, Southeast, Northeast) in the U.S.

The occurrence of HABs is a worldwide phenomenon. Several studies have suggested that climate change may be impacting the frequency and severity of harmful algal blooms (HABs) (Wells et al., 2020). Climate change could increase surface water temperature, cause more variable stratification (Wells et al., 2020), and thereby intensify the occurrence of cyanoHABs. A recent satellite remote sensing-based study of 71 lakes with surface area >100 km^2 distributed globally found that 68% had a significant increase in peak summertime bloom intensity from 1982 to 2012. In contrast, peak summertime bloom intensity decreased in 8% of the lakes studied (Ho et al., 2019). However, the study did not find any consistent relationship between the increase in bloom intensity and commonly reported co-variates of the algal bloom – temperature, precipitation, and fertilizer use in the surrounding watershed. In contrast, Wilkinson et al. (2022) conducted a study using 10-42 years of field-measured chlorophyll-a (chl-a; mg m^{-3}) data and reported no widespread algal bloom intensification in 323 lakes across American Midwestern and Northeastern states (Wilkinson et al., 2022). 10.8% of the water bodies had significant increases in bloom intensity, and 16.4% had significant decreasing trends (Wilkinson et al., 2022). Hou et al. (2022) analyzed Landsat satellite images from 1982 and 2019 and reported changes in lacustrine bloom occurrence by decades in 21,878 lakes spread across six continents. Their study showed that bloom risk increased globally in the 2010s decade except for Oceania. Previous satellite-based studies (Ho et al., 2019; Hou et al., 2022) have used the Landsat datasets, which have a better spatial resolutions (30 meters) and a reduced revisit frequency (one image every 16 days), especially considering only the cloud-free days over the bloom season. On the other hand, the Medium Resolution Imaging Spectrometer (MERIS) from Envisat and Ocean and Land Colour Instrument (OLCI) on Sentinel-3A and 3B dataset provides a moderate spatial resolution (300m) but frequent temporal coverage (one image every other day) to observe cyanoHABs. Moreover, the MERIS/OLCI sensors have a 620 nm band used to identify and confirm the presence of cyanobacteria, which is not available on Landsat. With the availability of data from MERIS and OLCI, we bridge the knowledge gap by more frequently monitoring cyanoHAB conditions with high fidelity in detecting cyanobacteria in the lakes across the CONUS.

In this study, our goal was to investigate the change in cyanoHAB biomass in a larger set of lakes across the CONUS by using data from the Cyanobacteria Assessment Network (CyAN) project (CyAN; Schaeffer et al., 2015) to assess how the bloom magnitude (Mishra et al., 2019) has changed in the OLCI era (2016-2020) relative to the MERIS era (2008-2011) in large lakes across the CONUS. The CyAN project has generated products from MERIS (2002-2012) and OLCI (2016-present)(Seegers et al., 2021). Datasets from the CyAN project have already been used for estimating areal extent (Schaeffer et al., 2022), temporal frequency (Clark et al., 2017; Coffer et al., 2020), occurrence (Coffer et al., 2020), and magnitude (Mishra et al., 2019). Here we expand beyond those studies by examining the combined MERIS and OLCI data sets to look at the spatial and temporal patterns in bloom magnitude and to identify environmental factors that may influence these patterns. In addition, we used several Land Use and Land Cover (LULC) datasets and physical data records such as precipitation and temperature to identify the critical LULC and physical factors contributing to the change in cyanoHAB magnitude. Specifically, we investigated 1) how the cyanoHAB magnitude has changed in the CONUS lakes over 2016-2020 compared to 2008-2011 and 2) what physical and LULC factors may have contributed to the change.

2. Materials and Methods

We used the remotely-sensed cyanobacteria bloom products termed the Cyanobacteria Index (CI_{cyano}), from the CyAN project (CyAN) to calculate cyanoHAB bloom magnitude in nominal chl-a units representing the annual mean cyanobacterial chl-a concentration in a lake over the recreational season. Then, we calculated the median bloom magnitude over 2008-2011 (four years) and 2016-2020 (five years) in 1881 lakes across the CONUS (Fig. S1, SM text 1) and used it for change analysis between those two observation periods using three different approaches described below. We then used Geographically Weighted Regression (GWR) to explain the spatial association of cyanoHAB bloom magnitude with physical factors related to temperature and precipitation and Land Use/Land Cover (LULC) surrounding the water bodies in three iterations - 1) the whole dataset, 2) a group of lakes where bloom magnitude had increased, and 3) a group of lakes where bloom magnitude has decreased. Then, we analyzed the distribution of physical and LULC covariates by lake groups, where bloom magnitude has increased or decreased, and reported if the difference in group medians is statistically meaningful using Cohen's d metric (Cohen, 1988; Sawilowsky, 2009). Finally, we linked the distribution of NOAA climate extreme index (CEI) and LULC variables over 2008-2011 (MERIS) and 2016-2020 (OLCI) with the change in bloom magnitude (Increase or Decrease). Specific details on data and methods are provided below. In addition, a conceptual workflow summarizing the data flow and analysis methods is provided in Fig. 1 for clarity.

2.1.Remote sensing data

Cyanobacteria Index (CI_{cyano})

The CI_{cyano} products were derived from 300 m resolution data from the MERIS sensor onboard the Envisat satellite for 2002-2011 and from the OLCI sensor on the Copernicus Sentinel-3A/3B mission for 2016-2020 through the CyAN project (CyAN). There is a temporal data gap in the satellite CI_{cyano} time series as a comparable sensor only became available in orbit when the MERIS replacement OLCI became operational mid-2016. While the MERIS sensor was intermittently available for CONUS from 2002 through 2007, continuous, full-resolution data were only available for CONUS between 2008 and 2012. The CI_{cyano} is calculated from the spectral surface reflectance ($\rho_s(\lambda)$; unitless). It is produced using l2gen, the NASA standard software packaged within SeaDAS (https://seadas.gsfc.nasa.gov) for processing Level-2 ocean color data, and projected to an Albers equal area projection. $\rho_s(\lambda)$ data are determined by removing Rayleigh radiances and gaseous transmission effects corrected for elevation from the instrument-observed top-of-atmosphere radiances, then converted to reflectance via normalization to downwelling irradiance at the sea surface (Seegers et al., 2021). Clouds are masked using a cloud detection algorithm (Wynne et al., 2018). Finally, adjacent pixels along each water body are masked to avoid land adjacency issues, including mixed land/water pixels, and to ensure the signals originating from land vegetation were identified and excluded from further analysis (Urquhart and Schaeffer, 2020). CI_{cyano} (Stumpf et al., 2016b; Wynne et al., 2008), was then calculated as follows.

$$\begin{gathered}
SS(681) = \rho_s(681) - \rho_s(665) - \{\rho_s(709) - \rho_s(665)\} * \frac{(681-665)}{(709-665)} \\
SS(665) = \rho_s(665) - \rho_s(620) - \{\rho_s(681) - \rho_s(620)\} * \frac{(665-620)}{(681-620)} \\
CI_{cyano} \begin{cases} = |SS(681)| & if\ SS(681)<0\ \&\ SS(665)>0 \\ = 0 & otherwise \end{cases}
\end{gathered} \tag{1}$$

Where $\rho_s(x)$ indicates Rayleigh-corrected surface reflectance measured at a band with a bandcenter of x nm. The algorithm is explained in greater detail elsewhere (Lunetta et al., 2015; Mishra et al., 2021; Mishra et al., 2019).

We applied the algorithm to both MERIS and OLCI which have equivalent bands by design. Our previous work has shown that OLCI requires a correction of 6% in the CI_{cyano} to match MERIS

CI_{cyano} (Wynne et al., 2021). While European Space Agency's OLCI calibration reprocessing is ongoing, we incorporated inter-calibration correction by multiplying OLCI CI_{cyano} with 1.06 to match the MERIS CI_{cyano} time series. The data sets were composited with the maximum CI_{cyano} value at each pixel for each sequential 7-day period for OLCI and MERIS starting in 2008. This approach reduces the impact of missing data due to clouds and underestimation of these blooms due to strong winds (Stumpf et al., 2012; Wynne et al., 2010). Less frequent coverage may miss more intense, especially scum-forming blooms if the only clear days during the composite were windy. Thus, the compositing process also minimizes the varying impact of wind on satellite-based cyanobacteria detection. Composite pixels with no valid data were excluded in the magnitude analysis, as described next.

2.2. Seasonal Bloom Magnitude

Cyanobacteria bloom magnitude is intended to represent the two key aspects of algal blooms: biomass quantity and bloom duration. Other metrics like frequency and spatial extent (Coffer et al., 2021; Schaeffer et al., 2022) provide information on temporal and spatial aspects of the bloom within a lake, but they do not address seasonal intensity. A spatial-temporal mean captures the quantity and duration of an entire lake over a season (or year). Accordingly, we estimated the bloom magnitude as spatiotemporal mean cyanobacteria biomass (Mishra et al., 2019) over the recreational season (May through October) within a lake as follows:

$$Mean\ bloom\ magnitude = \frac{a_p}{A_{lake}} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{T} \sum_{t=1}^{T} \sum_{p=1}^{P} CI_{cyano,p,t,m} \qquad (2)$$

The indices *P* and *T* in Eq. (2) represent the number of valid pixels in a lake or water body and the number of composite (time) sequences in each month (e.g., four in a month), respectively. *M* is the number of months in a season or annual study period; a_p is the area of a pixel, and A_{lake} is the area of the lake taken from the National Hydrography Dataset Plus version 2.0 (NHDPlusV2) lake vector layer (McKay et al., 2012) (see *SM text 1*). Using only valid pixel area to calculate spatial mean could add bias to the estimates. While more invalid pixels over high-concentration bloom events will underestimate, more invalid pixels over bloom-absence or non-detect pixels will overestimate the bloom magnitude. Therefore, we used the lake area in Eq. (2), which may introduce a systematic bias that could underestimate the results. As MERIS has a somewhat higher rate of invalid data, MERIS bloom magnitudes may be underestimated slightly more than OLCI. (The significance will be covered in the discussion.) Bloom phenology could vary slightly from southern to northern CONUS due to the seasonality in temperature and diurnal light availability. Additionally, snow/ice cover during winter is another significant issue in the northern CONUS. Therefore, in the high-latitude regions in the CONUS, we needed to exclude winter months. However, that would introduce positive bias in data quantity in the southern CONUS in the analysis. Therefore, we decided to use the recreational season as the time range for this study. Previous research has shown that the uncertainty in CI_{cyano} products is about $1x10^{-4}$ CI_{cyano} (Stumpf et al., 2016a). Therefore, we excluded all pixels < 1×10^{-4} CI_{cyano}. As CI_{cyano} values are relative index, we presented the spatio-temporal mean cyanobacteria bloom magnitude as nominal cyanobacterial chl-a concentration based on the relationship available for the CONUS lakes (Seegers et al., 2021).

$$Chl\text{-}a\ (mg\ m^{-3}) = 6620\ \times mean\ bloom\ \ magnitude\ \left(CI_{cyano}\right)\ \ (3)$$

The intercept term from Seegers et al. was not included as it was not meaningfully different from zero (Seegers et al., 2021). The slope term had an uncertainty of about 10%, which does not impact the analysis, as our computations are based on the CI_{cyano}, with chl-a used only for reporting. From here onwards, we refer to spatio-temporal mean cyanobacteria bloom magnitude as “bloom magnitude” for brevity.

2.3. Change analysis

A single change analysis was limited in demonstrating the various aspects of the change, such as increasing or decreasing temporal patterns, the difference in the size of the bloom magnitude, and proportional change between two time periods. In addition, there is a temporal data gap in the CI_{cyano} time series from 2012-2015. Therefore, we analyzed the change in bloom magnitude in the 2016-2020 period compared to the 2008-2011 period through (1) year-over-year change rate, (2) change between WHO alert levels, and (3) ratios of bloom magnitude between time periods.

2.3.1. Change rate in year-over-year bloom magnitude

We used Theil-Sen’s slope estimator (Sen, 1968) to assess temporal change patterns in the bloom magnitudes over the MERIS-OLCI study period (2008-2020). We also used Kendall’s τ (Kendall, 1938) for the Sen slope’s strength. Theil-Sen’s estimator for slope makes no assumptions about data and error distribution and provides an unbiased estimate of trend (Hirsch and Slack, 1984). Theil-Sen’s slope was expressed in the units of mg m^{-3} yr^{-1}. Kendall’s τ is a non-parametric statistical measure of rank correlation and is used to measure the ordinal association between two quantities. The value of the coefficient could vary from 1 when the ranking of the two measures is the same (perfect agreement) to -1 when the order of the two measures is reversed (perfect disagreement). |τ| values of < 0.3, 0.3-0.5, and > 0.5 are interpreted as weak, moderate, and strong strength in the relationship. Additionally, we determined the uncertainty in Sen slope estimates by converting CI_{cyano} uncertainty to a nominal chl-a. The detection threshold of CI_{cyano} is about $1x10^{-4}$ CI_{cyano} (Stumpf et al., 2016a). We assumed that a change of 1.324 mg m^{-3} of chl-a (2×10^{-4} $CI_{cyano}\times6620$) from 2008 to 2020 (13 years) cannot be measured due to the uncertainty associated with the retrievals. A difference of twice the uncertainty would conservatively accommodate uncertainty in the change analysis. Thus, we used a slope of 0.1 mg m^{-3} yr^{-1} (1.324 mg m^{-3} /13 years) as uncertainty in the change rate analysis. This value may appear small because it reflects the bloom as averaged over the lake and season, but it excludes any trend resulting from random patterns in the noise.

2.3.2. Change between WHO alert levels

We used the satellite-derived median bloom magnitude from the two periods to determine WHO alert levels (Chorus and Welker, 2021) for a given lake. WHO alert levels: Vigilance (chl-a of 3-12 mg m^{-3}), Alert Level-1 (chl-a of 12-24 mg m^{-3}), Alert Level-2 (chl-a > 24 mg m^{-3}) are monitoring and management action sequences that replaced the previous WHO guidelines of low, moderate, and high-risk categories for cyanoHAB monitoring (see *SM text 2*). In the current context, the alert level would indicate a lake's seasonal average alert level over the corresponding time period. Further, we highlighted when lakes changed alert levels. To capture the changes, we used a code that concatenates the 2008-2011 alert level, then the 2016-2020 alert level. E.g., code A1V represents a lake changed from Alert level 1 (A1) in 2008-2011 to vigilance (V) level during 2016-2020.

2.3.3. Bloom magnitude ratio

We took the ratio of the median annual bloom magnitude from the MERIS period (2008-2011) to the median bloom magnitude from the OLCI period (2016-2020). We expressed the ratio as a fold change. OLCI: MERIS ratio of <1, 1, and >1 indicates a decrease, no change, and an increase in bloom magnitude. We used $\log_2$ of the fold change to show proportional change in both positive (increase) and negative (decrease) directions more intuitively. With $\log_2$ of the ratio, a two-, four-, or eight-fold increase in magnitude equals a $\log_2$ fold change of 1, 2, or 3. An equivalent decrease (two-, four, or eight-fold, or 1/2, 3/4, or 7/8, respectively) would be expressed as a $\log_2$ fold change of -1, -2, or -3. Log_2 ratio value of 0 indicate no change between MERIS and OLCI. As the detection threshold is about $1\text{x}10^{-4}$ CI_{cyano} (Stumpf et al., 2016a), a difference of twice that ($2\text{x}10^{-4}$ CI_{cyano}) would conservatively accommodate uncertainty in the change analysis. In chl-a units (Equation 3), this value equates to 1.324 mg m^{-3}. We used a conservative estimate of ±2 mg m^{-3} as a threshold for identifying changes of higher confidence.

For further analysis, we grouped lakes into two categories. 1) *Increase*, where $\log_2$ OLCI: MERIS bloom magnitude was ≥ 1, and 2) *Decrease*: where Log_2 OLCI: MERIS bloom magnitude was ≤ -1.

2.3.4. Finding consensus among three change analyses

We used a majority voting approach to combine the change outcomes from three analysis methods and find a consensus. We chose this approach because it is straight-forward and as effective as other complicated schemes (Lam and Suen, 1997). Majority voting takes decisions from multiple classifiers or, in our case, change analysis methods and finds the most frequent output as the consensus. In this study, consensus occurs when the majority of the methods agree on the type of change. A lack of consensus would mean the observed change is uncertain. Using this approach, we can identify the set of lakes where the change outcomes have

1) a unanimous agreement (all three have the exact change outcome),
2) a majority agreement (two out of three have the exact change outcome), or
3) no agreement at all (all three have different change outcomes).

2.4.Climate data

We used monthly climate data to find correspondence between the observed differences in bloom magnitude and the climate variables. We downloaded monthly climate data aggregated within U.S. climate division boundaries from NOAA National Climate Prediction Center (NCPC) (NOAA-NCPC). The dataset included temperature (°F), precipitations (inch), and degree days (°F) data. Further, we derived additional features from the monthly climate data by taking the statistical mean, min, and max of a climate variable over a specific time period (a month or over several months), which included the maximum temperature from March to October (°C) or the sum of precipitation over May and July as cumulative precipitation (May-July) (mm). Although aggregation of climate variables was based on lacustrine cyanobacterial algal bloom phenology in the CONUS lakes (Coffer et al., 2020), final climate variables were not selected *a priori*. The variable selection process was entirely data-driven based on the Random Forest model to determine variable importance. In addition, we downloaded U.S. Climate Extreme Index (CEI) dataset for the warm season period (April-September) by climate region from the National Climate Data Center (NCDC) website (NCDC-NOAA). CEI quantifies observed changes in climate within the CONUS by summarizing a complex set of multidimensional climate variables in the U.S. within nine climate regions defined by the

National Center for Environmental Information (Karl and Koss, 1984). We used CEI for the observation period to find correspondence between a simplified and summarized state of climate and the cyanoHAB occurrences in the CONUS lakes.

2.5.Land Use and Land Cover (LULC) data

We downloaded annual LULC data for the years 2008-2011 and 2016-2020 from the United States Department of Agriculture (USDA) National Agricultural Statistical Service (NASS) website (NASS-USDA). For each lake, we extracted the corresponding LULC data within hydrological units at three hierarchical level that encloses the lake. Hydrologic Unit Code (HUC) is a hierarchical land area classification system created by the United States Geological Survey (USGS) based on surface hydrologic features in a standard, uniform geographical framework (HUC-USGS; Seaber et al., 1987). The United States is divided into successively smaller hydrologic units, which were classified into regions (HUC-2), subregions (HUC-4), basins (HUC-6), sub-basins (HUC-8), watersheds (HUC-10), and sub-watersheds (HUC-12). In this study, we used HUC-8, -10, and -12 to account for LULC and physical factors surrounding a lake at sub-basin, watershed to sub-watershed scale, and their effect on bloom magnitude. We extracted annual acreage information of relevant LULC types that included cropland area, wetland, grassland and pasture, forest and shrubland, and developed area within three Hydrological Unit (HU) boundaries with HU codes eight, ten, and twelve (HUC8, HUC10, HUC12) by converting extracted pixel counts from the cropland Data Layers (CDL) to acreage by LULC type. Further, we calculated the fraction of acreage of each LULC class in each HU by taking the area of the corresponding HU into account. We included HUCs at three different scales enclosing a lake (HUC8, HUC10, HUC12) and allowed the Random Forest feature selection (described below) to determine the dependence between the spatial scale of LULC variables and how they affect the bloom magnitude.

2.6.Feature selection with Random Forest model

We derived 146 input variables - 67 physical/climate variables summarized by the climate region (associated geographically) and 79 LULC variables at three hydrologic units enclosing the lakes. Considering many input variables, we used Random Forest (RF) regression model as a tool for feature selection. RF models have been effectively used to eliminate unimportant variables or features, and it has been instrumental even in datasets with a higher number of features (Chen et al., 2020). Based on feature rank and their importance, we selected eight LULC and climate features for modeling bloom magnitude, which are listed in Table 1. See *SM text* 3 for additional details about the RF model and selected features.

2.7.Geographically weighted regression (GWR)

GWR is a spatial statistical method for modeling spatially heterogeneous processes that allow the relationships between a response and a set of covariates to vary across geographic space (Brunsdon et al., 1996; Fotheringham et al., 1997; Fotheringham et al., 2001). GWR is a better approach (Kang et al., 2023) compared to classical linear regression when the effects of independent variables are not static over space. The key assumption in linear regression is that the data comes from an independent and identically distributed population of random variables. It does not assume that regression parameters in the model had relations with the geographical location of variables. However, GWR incorporates spatial information into the regression model, allowing uncovering of the spatial variation in the relationship among variables.

We used GWR in this study to model localized physical and anthropogenic factors surrounding a lake, listed in Table 1, and their association with the bloom magnitude in a lake. The primary component of GWR is the spatial weight matrix in which closer observations are assigned larger weights defined by spatial kernel functions such as a Gaussian function (Brunsdon et al., 2002).Thus, localized regression models are calibrated by data from surrounding locations. GWR calibrates *n* number of regression models, where *n* is the number of lakes, producing *n* sets of model coefficients and model R^2 (local R^2), which can be visualized with descriptive statistics or as a surface map. We scaled the eight independent variables such that they vary from zero to one before training the GWR regression models. Therefore, we can compare the model coefficient maps and the relative effects of the independent variables based on the magnitude or size of the coefficients. Additional mathematical details of GWR are available in *SM text 4*. Note that we didn't select the variables "locally"; instead, we selected the variables 'globally' using a Random Forest model. We wanted to capture local relations. However, we didn't want to train over-fitted GWR models that can happen due to local variable selection. Additionally, we tried to select meaningful variables with broader significance across the CONUS to draw meaningful conclusions in a CONUS-wide study.

3. Results

3.1.Change in Bloom magnitude

3.1.1. Temporal change rate

In 1881 largest lakes across CONUS, bloom magnitude was lower over the OLCI period (2016-2020) than the last four years of the MERIS period (2008-2011) (year-over-year median range for OLCI of 0.8 – 1.1 mg m^{-3} vs MERIS of 1.4 – 1.7 mg m^{-3}) (Fig. S2, Table S1). A widespread decrease in bloom magnitude from the MERIS (2008-2011) period to the OLCI (2016-2020) period was observed in lakes across the CONUS. Of the 1881 lakes, the Sen slope, a statistically robust metric for analyzing change over time in time series data (Hirsch and Slack, 1984), was negative in 1447 lakes (77%). Sen slope was positive in only 434 lakes (23%). However, the lake counts with decreasing and increasing pattern reduced to 415 (22%, Kendall's τ of $\leq$ -0.3 and Sen slope < 0.1) and 135 (7%, Kendall's $\tau \geq 0.3$ and Sen slope > 0.1), respectively, when Kendal's τ and Sen slope uncertainty were used for assessing the strength of the change (Fig. 2a, Fig. 3a-b). Although a more decreasing than increasing change was observed, the Slope's strength, per Kendal's τ, was weak in majority of the lakes (n=1377, 73%) (Fig. 3c). Of 1377 lakes with Kendal's $|\tau| \geq 0.3$, 413 lakes had extremely small Sen slopes that fell within the uncertainty band of - 0.1 to 0.1 mg m^{-3} yr^{-1} (see *Methods* for uncertainty calculation). Similar changes were observed when bloom magnitude over 2003-2011 was used, underlining that the observed temporal change patterns were valid starting in 2003 (Fig. S3, *SM text 5*).

3.1.2. Change between WHO alert levels

Most lakes were below the WHO Vigilance (V) category, we called it No-risk (N), over the observation periods. During 2008-2011, 1130, 434, 195, and 122 lakes were in no-risk, vigilance (V), alert level-1 (A1), and alert level-2 (A2) categories, respectively (Figs. 2b). In the 2016-2020 period, the number of lakes in the no-risk category increased to 1299 (+15%), while lake counts in the V, A1, and A2 categories decreased to 389 (-10%), 140 (-28%), and 53 (-56%), respectively. More lakes (403, or 21%) changed to a lower WHO category (Fig. 2c, green highlighted bars) than the number of lakes (70 or 3.7%) moving to a higher category (Fig. 4, Fig.

2c, red highlighted bars). The shift of lakes from V to No Risk level (N, < 3 mg m^{-3} chl-a) and A1 to V categories contributed to the significant decrease in the bloom conditions. On the other hand, 35 and 21 lakes from the N and V categories moved to V and A1 categories, highlighting the bloom magnitude increase in those lakes in recent years. However, 75% of the lakes (*n*=1408) maintained the same WHO category over the study period, out of which 1093 lakes were at no risk level over both time periods. As expected, of the 413 lakes with extremely low Sen slopes within the uncertainty band of - 0.1 to 0.1 mg m^{-3} yr^{-1}, 411 (99%) of them fell within the NN (No-risk during both periods) category (Fig. 2a, c).

3.1.3. Median bloom magnitude ratios

The decreasing pattern was even more compelling when we summarized the change by the ratio of median bloom magnitudes from OLCI and MERIS periods (Fig. 5). 83.3% of the lakes decreased in bloom magnitude in 2016-2020 compared to 2008-2011 (Figs 5 and 6a). Only 312 lakes (16.7%) had an increase in bloom magnitude. However, when accounted for uncertainty in the change analysis (|chl-a difference| > 2 mg m^{-3}), 27% of lakes were identified where bloom magnitude decreased. Of 27% of lakes, there were 11.1% where bloom magnitude decreased up to 50% (log2 (OLCI: MERIS ratio) of -1 to 0) and another 11.3% of lakes where magnitude decreased 50-75% ($\log_2$ (OLCI: MERIS ratio) of -2 to -1) (Fig. 6b). The other 5% had a decrease of more than 75% ($\log_2$ (OLCI: MERIS ratio) < -2) of bloom magnitude from the MERIS period. Conversely, when uncertainty in data and methods are considered, bloom magnitude increased in only 5% of the lakes (Fig 6a). In that group, bloom magnitude increased 1-2-fold in the majority of the lakes (n=56, 3%), and greater than 2-fold in 2.07% of the lakes (Fig. 6b).

3.1.4. Consensus among change analyses

The three different analyses showed consistency in change or no change (Fig. 7, left panel). 74% of the lakes had the same result in all three change analysis methods (unanimous consensus) (NNN, DDD, and III counts in Fig 7, right panel). None of the lakes showed an 'Increase' in one method and a 'Decrease' in another, indicating consistency among these methods. While 71% of the lakes showed no change based on the majority of methods, 25% of the lakes had a decrease, and only 4% had an increase (Fig. 7, right panel). Of 1335 lakes in the (majority consensus) 'No Change' category, bloom magnitude decreased in most of them when uncertainty in data would not be considered, based on temporal change rate (n=989, 74%) and bloom magnitude ratio method (n=1104, 83%). Thus, the bloom magnitude in most of the lakes in the 'No Change' category either decreased or the observed change was in the uncertainty range.

3.2.LULC and physical factors

Model coefficients from a GWR analysis highlight covariates' non-stationary effects (effects that vary over space) on the bloom magnitude across CONUS, which is evident in the model coefficient surface maps (Fig. S4)**.** The model's performance in terms of R^2 (median R^2 = 0.46, 3rd quantile R^2 =0.58) across the CONUS implies there were effects from local anthropogenic and natural processes on bloom magnitude (Table 2, extended Table S2). The fraction of grassland and pasture in HUC10 and crop acreage in HUC12 are the top local factors (in the GWR neighborhood, see *SM text 4*) based on the size of the median parameter estimates (Table 2). For 50% of the lakes, a higher proportion of grassland and pasture acreage and crop acreage in the nearby hydrologic units (HUCs) are positively associated with higher bloom magnitude. The impact of grassland and pasture on bloom magnitude is predominantly positive along the west coast, along the Mississippi River delta, eastern Texas, and northern Michigan region, and primarily negative in central Texas, Minnesota and Wisconsin, Central Florida, Ohio River valley, and in the North Carolina coastal area (Fig. S4). Similarly, crop acreage fraction positively affected the bloom magnitude in the West North Central, Northwest, and Southwest climate region and negatively associated in the west coast, East North Central, and South climate regions, Florida, and Maine.

Maximum temperature from May to October, T_{max} (May-Oct), and cumulative degree days from May to October, CDD (May-Oct), are the top climatic variables associated with the bloom magnitude, with associations to half of the lakes across the CONUS (Table 2). T_{max} (May-Oct) was positively associated with bloom magnitude in the Central, South, and Southeast climate regions. T_{max} (May-Oct) was negatively associated in Florida and the southern tip of Texas, possibly suggesting high-temperature stress. Bloom magnitude in lakes in the Northeast and East Northcentral climate regions (New England region, Michigan, Wisconsin, and Minnesota) was associated negatively with T_{max} (May-Oct). Spatial patterns of Cumulative CDD (Mar-Oct) effect on bloom magnitude is inverse of T_{max} (May-Oct) coefficient surface (or inverse relationship with bloom magnitude) with exceptions in central Florida and part of the Northeast climate region. Although cumulative CDD (Mar-Oct) and T_{max} (May-Oct) capture similar environmental information (temperature), their association with bloom magnitude is the opposite of each other.

We also analyzed the '*Increase*' and '*Decrease*' groups based on bloom magnitude ratio (see section 2.3.3) of lakes and their corresponding covariates to see any association with bloom magnitude (Fig. S5). The median $PDSI_{AN}$ (%) in the '*Increase*' group (median: 6% of the area) was lower than the '*Decrease*' group (median: 23% of the area), with a medium to large difference between the group means based on Cohen's *d* (Cohen, 1988; Sawilowsky, 2009) (d =-0.6, Table S3, Fig. S5). In other words, if PDSI was above normal in a larger fraction of a climate region area, bloom magnitude in lakes within that climate region decreased over the 2016-2020 period compared to 2008-2011. Similarly, in lakes where bloom magnitude has doubled, 50% of the lakes (median) have experienced ~76 mm less cumulative precipitation during June-July of the corresponding year than the lakes in the '*Decrease*' group. The difference between the group means was large, based on Cohen's *d* (d =-0.8, Table S3). Thus, lower cumulative precipitation and lower percent of the area with PDSI above normal conditions in a climate region were associated with an increase in bloom magnitude (Fig. S5, Table S3). On the other hand, the median T_{max} (May-Oct) in the '*Increase*' group was larger (2.5 °C) than in the '*Decrease*' group. Thus, the difference in T_{max} (May-Oct) between the two groups was of medium strength (per Cohen's $d =0.49$, Fig. S5, Table S3). Median cumulative CDD was 171.5 °F higher in the '*Increase*' group (Table S3). The differences between the two groups for cumulative CDD were of small strength (per Cohen's $d = 0.2$, Table S3).

3.3.The U.S. Climate Extremes Index (CEI) and bloom magnitude spatiotemporal patterns

The spatial pattern of decrease in bloom magnitude was prominent in West North Central, South, Southeast, and Central climate regions, where ~20-40% of the lakes experienced a decrease in cyanoHAB magnitude per majority change among three methods (Fig. 8). Over the two observation periods, the regional patterns in the CyanoHAB decrease (Fig. 8) are similar to the patterns of the $PDSI_{AN}$ (Fig. 9a). Over the recent years, 2016-2020, PDSI was above normal in a larger area in the above climate regions (excluding West North Central) compared to the 2008-2020 period (Fig. 9a). In contrast, PDSI was below normal in significantly larger parts over the 2008-2011 period highlighting dryer warm season conditions (Apr-Sept) in the eastern CONUS (Fig. 9b). Similarly, the maximum air temperature was above normal in a more significant part of those climate regions over the 2008-2011 period (Fig. 9c).

Moreover, eastern CONUS, excluding the northeast, received up to ~137 mm of more cumulative precipitation and days of precipitation above normal in a larger part of the region in recent years (Fig. 9d-e, S6a). Similarly, one-day precipitation in parts of the West North Central climate region was higher over 2016-2020 compared to 2008-2011 (Fig. 9f). Increased precipitation in the U.S.-Midwest and Northeast lakes has been linked with decreasing trends in bloom magnitude (Wilkinson et al., 2022) . Therefore, above-normal conditions in wetness across a more prominent part and normal or below-normal conditions in maximum temperature over the warm season may have caused the recent decrease in bloom magnitude in the eastern part of the CONUS.

In comparison to other parts of the CONUS, the western part saw the highest proportion of lakes (6-12%) with an increase in CyanoHAB magnitude (Fig. 5, Fig. 8). That could be due to the fact that up to 10-95% of the area in the western U.S. experienced above-normal maximum temperature and below-normal days of precipitation. Similarly, median T_{max} (May-Oct) over 2016-2020 was up to ~2°C higher than the 2008-2011 period in the western half of the CONUS (Fig. S6b). Rising temperatures favor cyanobacteria as they grow better at higher temperatures (Havens and Paerl, 2015; Paerl and Huisman, 2008). Accordingly, the temperature is expected to increase the occurrence and magnitude of freshwater cyanobacteria (Wells et al., 2020). In addition, we observed ~10% of increased usage of land for growing row crops and all other crops in the West North Central climate region (Figure S7). In recent years, increased crop production may have contributed to more nutrient loading in the watersheds. An increase in corn acreage (a row crop) has been linked with an increase in nutrient loading to Lake Erie (Michalak et al., 2013). Thus, warmer conditions and more frequent above-normal precipitation days in the West North Central climate region may have caused increased nutrient loading that may have increased bloom magnitude.

4. Discussion

Our results provide empirical evidence of a recent decrease in bloom magnitude over 2016-2020 compared to 2008-2011. All regions of the CONUS had more lakes that showed a decrease in bloom magnitude than an increase, even including areas where the maximum temperature increased (Fig. 8). Western U.S. did have the greatest increase in temperature, and compared to other regions, it had a higher proportion of lakes with an increase in bloom magnitude. Several studies have linked the warming of surface water to an intensification of algal blooms and postulated the future widespread intensification of algal blooms with the changing climate (Gobler, 2020; Paerl, 1988). (Gobler, 2020; Paerl, 1988). While these changes may cause intensification in the long term, the recent record from this study does not show such a pattern over the study period. A similar study in the U.S. Northeast and Midwest, including data from the 1980s to 2010s, found that only 10.8% of the 300 lakes experienced algal bloom intensification and concluded no widespread intensification in bloom intensity (Wilkinson et al., 2022) . We found similar results for the equivalent regions (East North Central, Central, Northeast), with negligible lakes seeing an increase (Fig. 8). Across the CONUS, only 4% of the lakes experienced a significant increase in bloom magnitude from 2008 to 2020. Although Wilkinson et al. had similar results, the observation periods were different: Wilkinson *et al.* used chl-a time series of varying observation periods (10 to 42 years with a median of 14 years) for different lakes, which can constrain comparisons between lakes. Nonetheless, our results and Wilkinson *et al.* indicate that a larger fraction of lakes decreased in bloom magnitude than increased.

Spatial patterns of recent change from this study are consistent with the CONUS part of the decadal study of lakes around the globe by Hou et al. (2022). Although Hou et al. focused on global lakes, we could compare the change patterns from their study to ours by selecting the CONUS area from their report that occurred during our observation time. From 2000-2010 to 2010-2019, they reported a decrease in bloom occurrence (or frequency of satellite-detected blooms) in the lakes they analyzed in the eastern U.S. and an increase in bloom occurrence in the western U.S. Thus, the long-term change in bloom occurrence can be region-specific; opposite patterns in temporal change are possible on a continental scale (Hou et al., 2022).

Recent (OLCI-based) temporal changes in cyanoHAB spatial extent in more than 2000 lakes across the CONUS were analyzed by Schaeffer et al. (2022). They found an increase from 2016 to 2020. We have similar results. We aggregated all CONUS data from OLCI (2016-2020) and found an average increase in bloom magnitude (0.25 mg m-3 yr-1), corresponding to the spatial extent increase observed by Schaeffer et al. (2022). However, this increase was much smaller than the decrease from 2008-2011 to 2016-2020, so OLCI is still well below the 2008-2011 bloom magnitude baseline. CyanoHAB magnitude changes varied dependent on the temporal scales considered, and we cannot assume that patterns over a few years represent longer trends.

LULC and climate covariates as predictors of bloom magnitude are consistent with other studies (Iiames et al., 2021; Myer et al., 2020). For example, out of 75 landscape and lake physiographic predictor variables considered by Iiames et al. (Iiames et al., 2021), percent area forest, percent evergreen forest, percent area row crop, and percent area evergreen forest were among the top-ten predictors. Myer et al. (Myer et al., 2020) reported that the important covariates are surface water temperature, ambient temperature, precipitation, and lake geomorphology. While CDD and $PDSI_{AN}$, which we found to be important, were not explicitly considered by Myer et al., they are related to their climate variables.

Here, we used data from two sensors to assess the change in bloom magnitude with the same algorithm. While the two sensors are not identical, OLCI was designed to be the continuity mission to MERIS with nearly-identical MERIS bands (ESA). MERIS calibration has been established through four iterations of processing (Ansko et al., 2015), while OLCI calibration is still being refined, necessitating the cross-calibration. Moreover, MERIS and OLCI have similar field-of-view (68.5°), comparable swath width (1150 km for MERIS and 1270 km for OLCI), and smile effects (1.7 nm for different cameras and 1.0 nm within one camera for MERIS and 1.4 nm for different cameras and 1.0 nm within one camera for OLCI) (D'alba and Colagrande, 2005; Vicent et al., 2016; Zurita-Milla et al., 2007). Moreover, previous work inter-calibrated the OLCI CI_{cyano} to match MERIS CI_{cyano} (Wynne et al., 2021), which was applied to our analysis. Therefore, the difference in CI_{cyano} from the two sensors would be minimal, with negligible effect on our analysis, as the observed changes (%) are several folds larger than the expected uncertainty in the cross-calibration (<0.5% within geographic regions) (Wynne et al., 2021). To avoid issues with possibly different minimum detection limits between the sensors, we excluded all pixels with CI_{cyano} values less than the uncertainty threshold of 1×10^{-4}. The compositing of maximum values over a 7-days period reduces the impact of winds on strongly buoyant (i.e., scum-forming) blooms (Wynne et al., 2021). However, this could still impact the analysis; higher frequency data collection during OLCI period will increase the likelihood of getting imagery on clear and low wind days (Wynne and Stumpf, 2015; Wynne et al., 2010). As OLCI has a higher frequency (two satellites, wider swath, angled to reduce glint), more blooms may be recovered, which could bias OLCI toward higher magnitudes over 2016-2020. However, while OLCI might return more data, MERIS may be underestimated because of the difference in retrieved data. (MERIS 2008-2011 has a 10% lower data return than OLCI. As a result, more lakes may have seen a decrease in significant bloom intensity than we reported. That is the opposite of the observed change – a decrease in magnitude from 2008-2011 to 2016-2020. Some lakes may have cyanobacteria that are below detection but of consequence. On the other hand, a standard water sample from a location near the shore where accumulation occurred would overstate the true magnitude of the bloom in the lake. Finally, the satellite-derived chl-a estimates have an uncertainty with 60% mean absolute error at the national scale (Seegers et al., 2021), 84% overall agreement against in-situ toxin data (Mishra et al., 2021), and 73% overall agreement with state-reported events (Whitman et al., 2022). However, the CI_{cyano}-chl-a algorithmic error reported by (Seegers et al., 2021) is within the previously reported possible uncertainty range of 39% to as high as 68% in the field chl-a measurements (Gregor and Maršálek, 2004; Trees et al., 1985). Moreover, World Health Organization (WHO) thresholds between alert levels have broader chl-a bands, hence greater uncertainties (Chorus and Welker, 2021). In addition, spatial-temporal representation from discrete samples does not reflect the larger systems observed by moderate-resolution satellite sensors. For example, discrete *in situ* water samples in cyanobacteria blooms may differ by as much as two orders of magnitude within tens of meters. Therefore, it is practically impossible to collect representative water samples when subsurface aggregations of cyanobacteria or surface scums occur (Kutser, 2004). Thus, the observed error could also be due to high variability and uncertainty in the field data. On the other hand, satellite-measured CI_{cyano} measurements have high temporal consistency. Most regional deviation from the national chl-a calibration would be systematic in each lake. For example, the standard error in the CI_{cyano}-chl-a slope, parameterized in several Southern Florida lakes, was ~7% (Tomlinson et al., 2016). Therefore, it would not significantly affect the change detection analysis as we compare how the bloom magnitude changed in the same lake over time.

5. Conclusion

Our study highlights the spatially varying interactions between cyanobacteria presence, LULC, and physical factors. Temporal changes in bloom occurrence can vary significantly at country, continental, and global scales (Hou et al., 2022), potentially due to the interaction between precipitation, temperature, and LULC. Moreover, temperature and precipitation do not monotonically increase across a continent in response to increases in CO_2. Therefore, spatiotemporal change patterns in HAB conditions should be assessed on relevant scales for better spatial granularity. While the CONUS had an overall recent decrease in bloom magnitude compared to 2008-2011, there were clear regional differences, with some regions showing no change or an increase. Moreover, bloom magnitude has been increasing since 2016 in seven of nine climate regions, excluding the Northeast and Southwest, where the change is negligible, highlighting the cyclicity in bloom magnitude, which may be due to the cyclicity in temperature (Li et al., 2021) and precipitation signals (Armal et al., 2018). We should also expect that temperature and precipitation cyclicity will continue, and some regions will see an increase in bloom magnitude over the next decade if climate patterns conducive to cyanobacteria growth occur. Similarly, changes in the landscape and land use can also alter the dynamics. For example, a change in fertilizer practice with no-till farming altered the bioavailable phosphorus in the Lake Erie watershed, leading to a greater susceptibility of the lake to cyanobacterial blooms in the last decade (Baker et al., 2014), which may be mitigated by additional changes in agricultural practice. Finally, as we saw regional patterns in the CONUS, we may not expect any systematic global patterns in response to climate. That is because several other factors, such as lake depth and morphology, nutrient level, and the surrounding landscape and hydrology, can affect the climate-bloom response interaction (Hou et al., 2022; Kosten et al., 2012; Qin et al., 2020). And certainly, the level of eutrophication will vary across countries and climatic zones. Therefore, extensive ecosystem-scale mechanistic modeling is required to quantify the impacts of increased temperature and nutrient loading on cyanoHABs at multiple spatial scales.

Figures

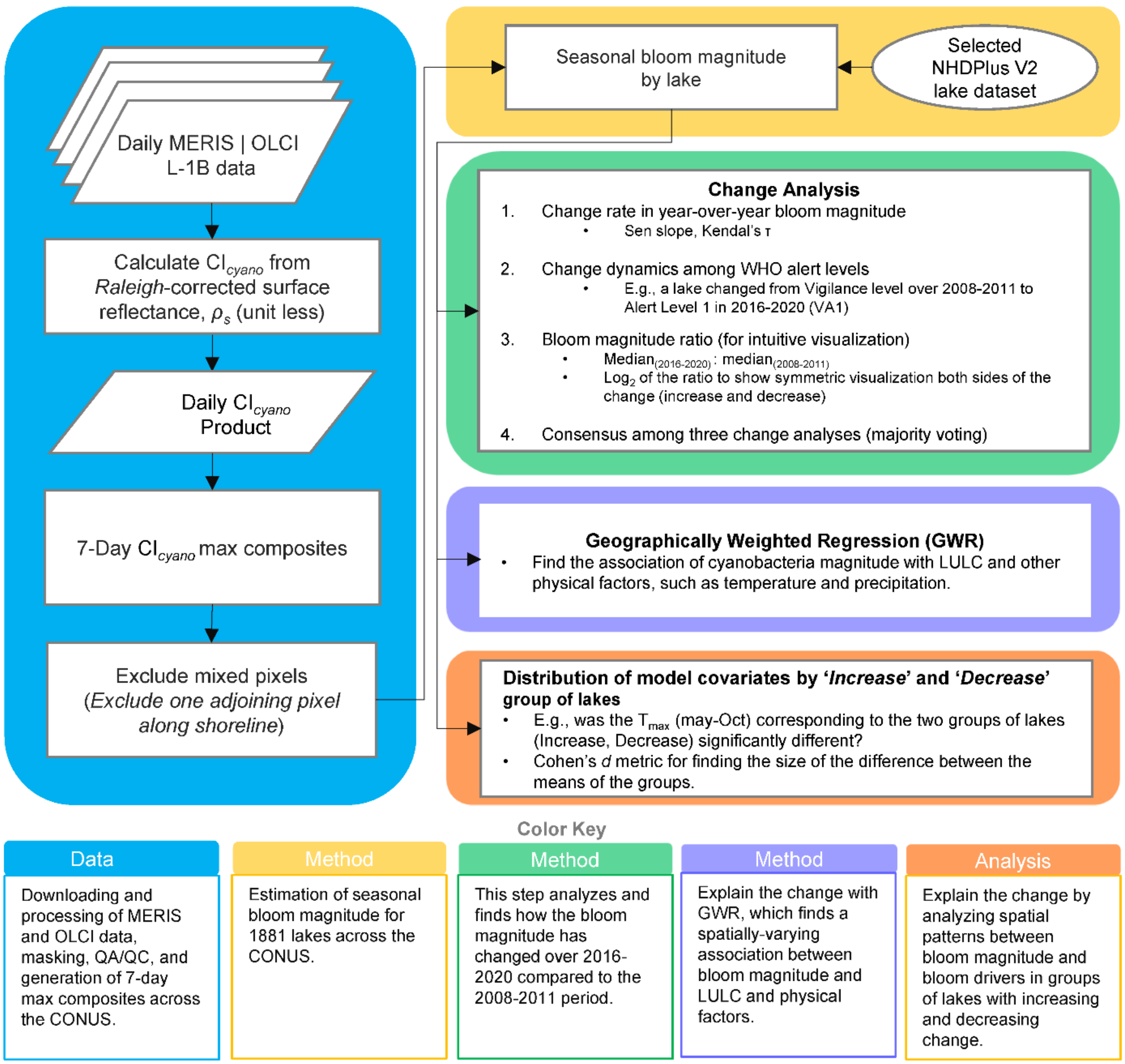

Figure 1. Satellite data processing and analysis workflow highlighting key methods and steps carried out to study how cyanobacteria bloom magnitude has changed in the CONUS lakes in 2016-2020 compared to 2008-2011.

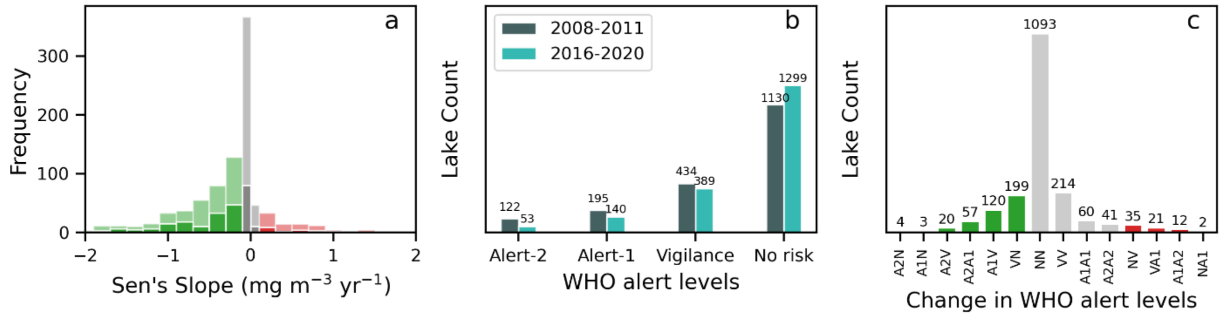

Figure 2 a) Distribution of Sen's slope from the bloom magnitude change rate analysis. Green and red colors highlight the negative and positive changes in the time series. Slopes that fell within the uncertainty range (-0.1 to 0.1 mg m^{-3} yr^{-1}) are highlighted in gray color. Lighter and darker shades of color indicate two significance levels - Kendall's $|\tau| \geq 0.3$ and ≥ 0.5, respectively. Sen slopes varied from -6.34 to 3.98 mg m^{-3} yr^{-1}, but the axis was truncated to highlight the majority of the distribution; b) the number of lakes in each World Health Organization's CyanoHAB alert level (Chorus and Welker, 2021) in the contiguous United States. Gray and turquoise bars indicate the data from MERIS (2008-2011) and OLCI (2016-2020) time periods; c) Number of lakes in each bloom status change class. Bar labels show the change from one alert level to another. E.g., A1V represents lakes changing from Alert level 1 in 2008-2011 to vigilance level during 2016-2020.

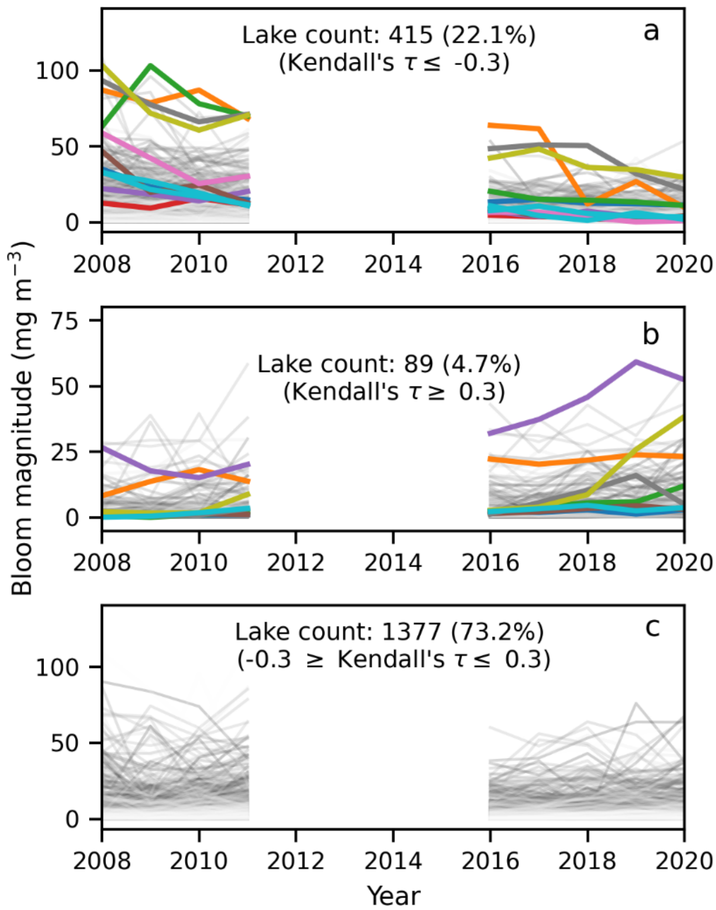

Figure 3. Cyanobacterial chl-a time series in lakes as observed from the satellite-derived data. a) lakes where the bloom magnitudes have moderately or strongly decreased; b) Lakes where bloom magnitudes have moderately or strongly increased; c) lakes with weak decreasing or increasing patterns over the observation period. Gray lines indicate change over time with moderate (Kendall's |τ| > 0.3), and colored lines indicate strong (Kendall's |τ| > 0.5). Note satellite observation gap from 2012 through 2015.

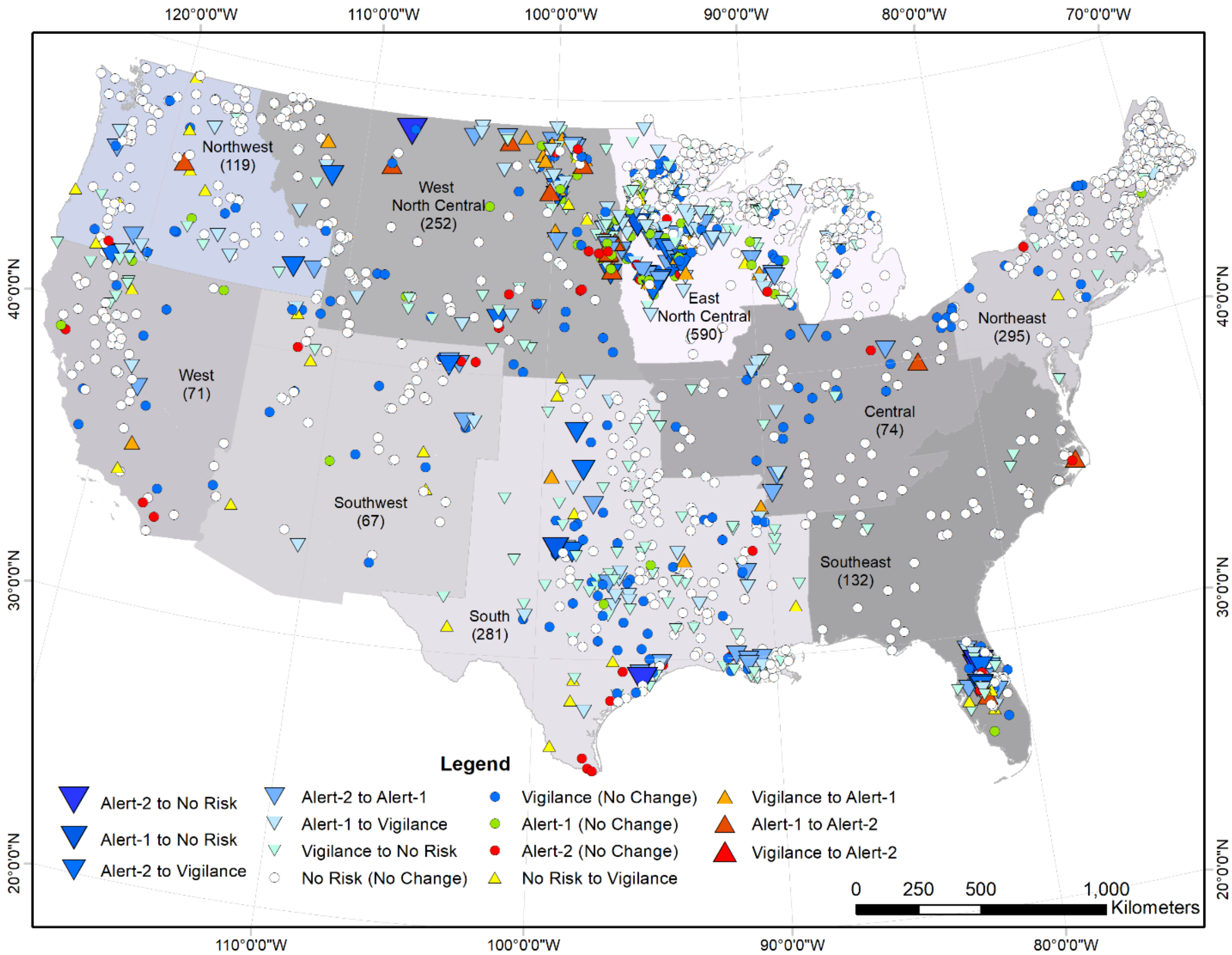

Figure 4. Change in cyanobacteria bloom magnitude as observed from MERIS (2008-2011) and OLCI (2016-2020) observations. Markers represent 1881 of the largest lakes in the contiguous United States that can be resolved with 300x300 m pixel resolution satellite data and have nine years of observation; their shapes show the bloom change among WHO alert levels (Chorus and Welker, 2021). As adopted from NOAA National Center for Environmental Information (NCEI) (Karl and Koss, 1984), nine climate regions are provided in the background for reference. In addition, lake counts in each climate region are provided as part of the labels.

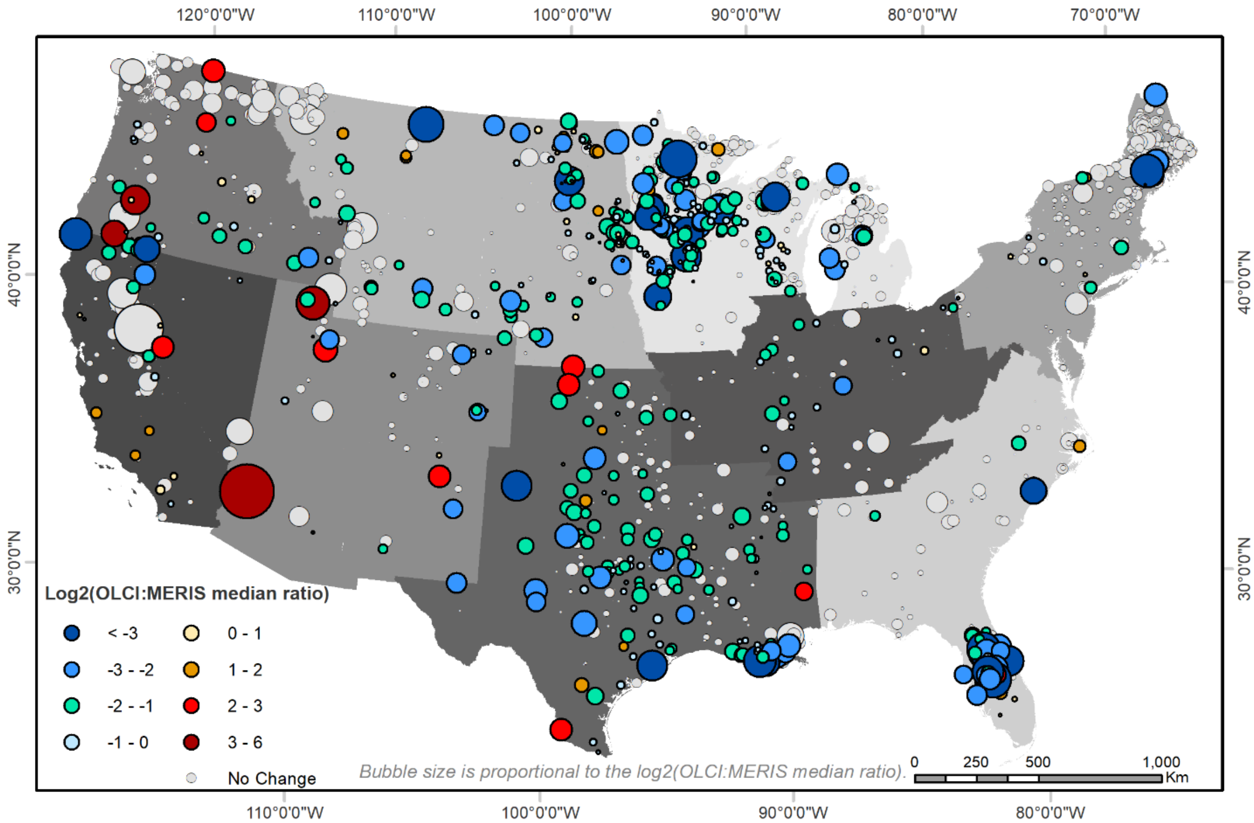

Figure 5. Changes in median bloom magnitudes in lakes between the two study periods 2008-2011 and 2016-2020. Cooler colors indicate a decrease in median bloom magnitude, and warmer colors indicate an increase. A $\log_2$ fold change of 1, 2, and 3 shows an increase in bloom magnitude of two-, four-, or eight-fold. Similarly, a $\log_2$ fold change of -1, -2, and -3 indicates halving (50% decrease), quartering (75% decrease), and 87% decrease. $\log_2$ (OLCI: MERIS ratios) of 0 indicate no change. Bubble size is proportional to $\log_2$ (OLCI: MERIS ratio). Gray bubbles highlight the lakes where the absolute difference between the magnitudes from the two study periods was ≤ 2 mg m^{-3} of chl-a.

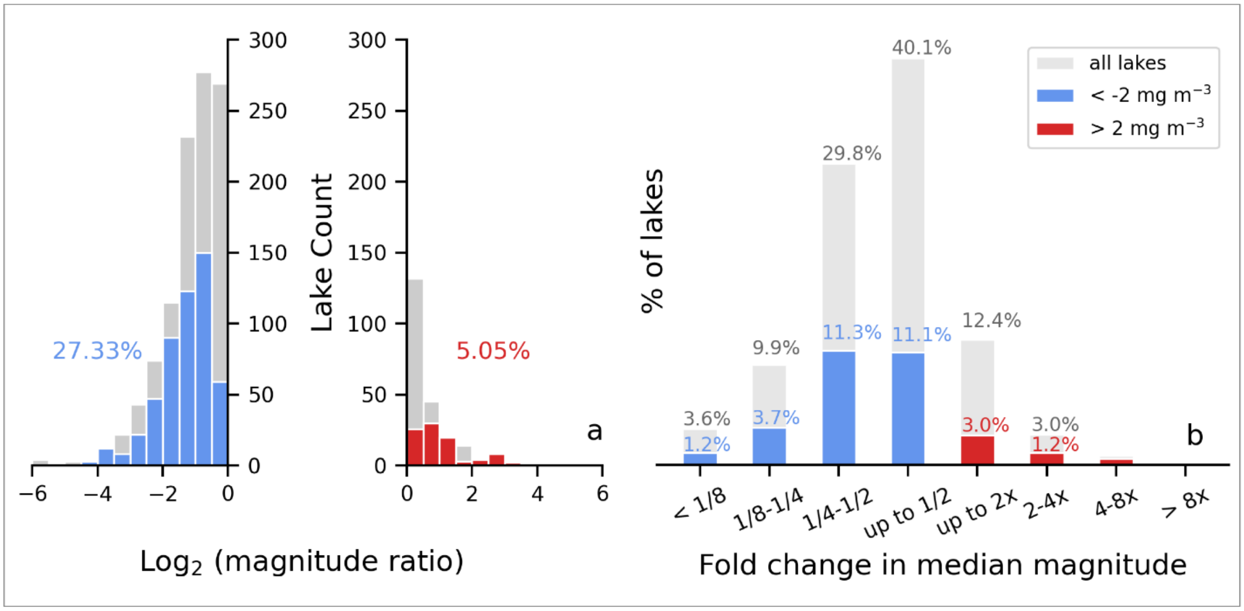

Figure 6. a) Distribution of bloom magnitude ratios across the CONUS. The histogram with blue (left) shows the lakes where bloom magnitude decreased, whereas the one with red (right) shows the ratio when the median magnitude over the OLCI period increased. A $\log_2$ fold change of 1, 2, and 3 shows an increase of 2-, 4-, or 8-fold. Similarly, a $\log_2$ fold change of -1 and -2 indicates a decrease of 1/2, 3/4, and 7/8. Log_2 (OLCI: MERIS ratios) of 0 indicate no change. The gray histogram represents lakes where the change in bloom magnitude fell within ± 2 mg m^{-3}, and b) Same data summarized as percent of lakes in increase/decrease discrete bins.

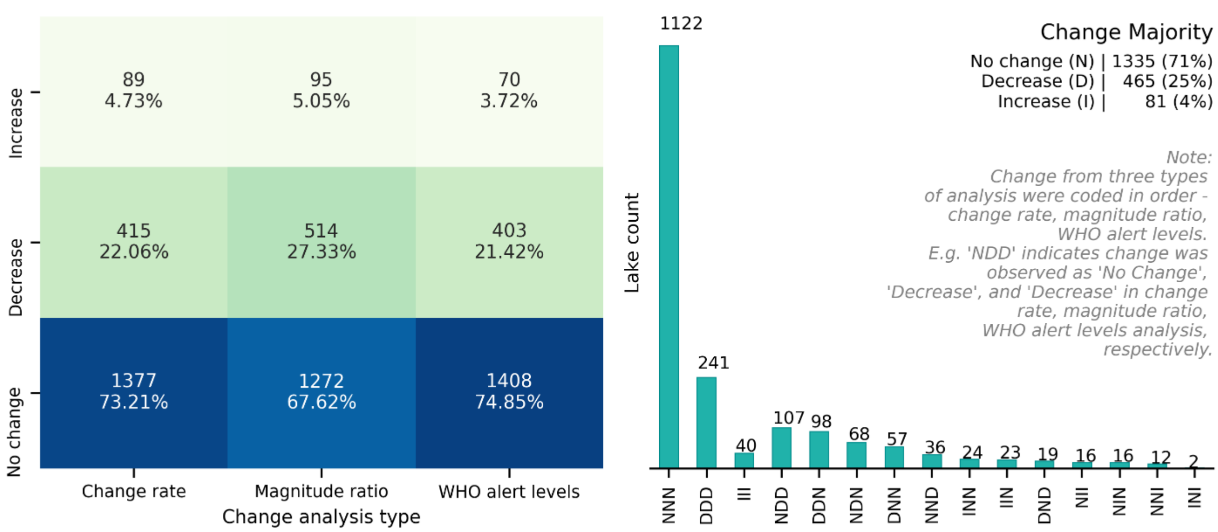

Figure 7. Left panel: Summary of change analysis from three different methods. Right panel: Consensus in change analysis as observed through three change analysis methods. Each bar represents the lake count with the change observed from year-over-year change rate, ratios of bloom magnitudes, and change between WHO alert levels. E.g., the bar labeled NDD represents that the change was observed as ‘No change’, ‘Decrease’, and ‘Decrease’ from year-over-year change rate, ratios of bloom magnitudes, and change between WHO alert levels, respectively.

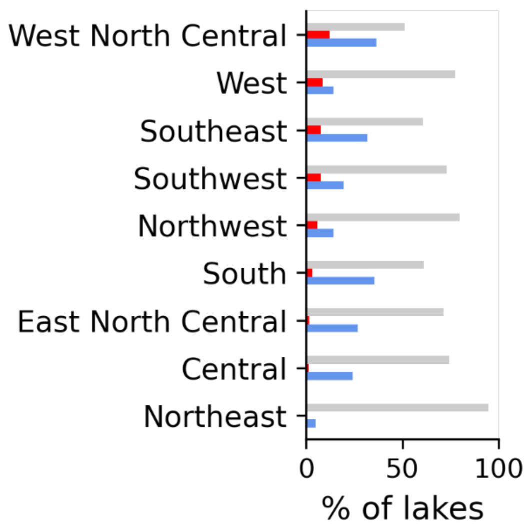

Figure. 8. Proportion of lakes experiencing an increase or decrease in bloom magnitude as observed by the change majority (change determined by two out of the three methods) in each climate region. Blue, red and gray bar colors indicate ‘Decrease’, Increase’, and ‘No change’ in bloom magnitude, respectively.

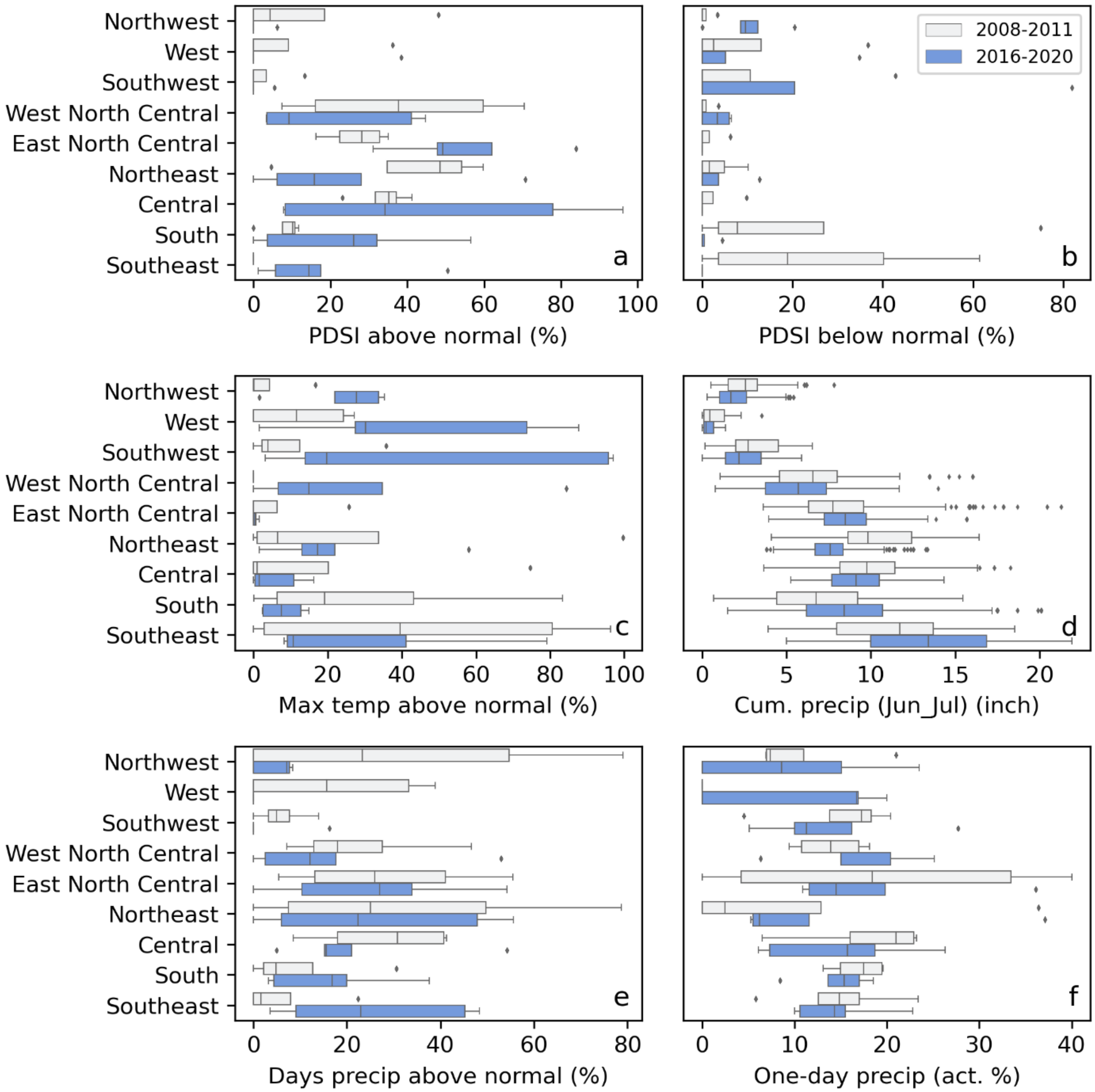

Figure 9. The distribution of NOAA climate extreme index (CEI) components over the warm season (Apr-Sep) in each climate region in the CONUS. Cumulative precipitation (Jun-Jul) is not a component of the CEI, but included here for comparison. Left and right bound of the boxes represent the first and third quartiles, respectively. The whiskers show 1.5 times of the interquartile range. The vertical bars in the middle of the boxes are the median, and the diamond markers are detected as outliers.

Table 1. List of selected land use and land cover (LULC) and climate features chosen by a Random Forest model and used in the Geographically Weighted Regression (GWR).

Selected features	**Description**
All_crops_acr_pct_hu12	Percentage of the total acreage of all croplands in the HUC 12, representing the agricultural activity in the hydrologic unit surrounding a lake under study.
Forest_shrub_acr_pct_hu8	Percent area of the HU with code eight surrounding a lake covered by forest and shrubland.
Grassland_pasture_acr_pct_hu10	Percent area of the HU with code ten surrounding a lake covered by grassland and pasture.
Wetland_acr_pct_hu12	Percent area of the HU with code 12 surrounding a lake covered by wetlands.
PDSI above normal ($PDSI_{AN}$)	Palmer Drought Severity Index (PDSI) is a standardized index computed from temperature and precipitation data to estimate relative dryness.
T_{max} (Mar-Oct) (°C)	Maximum temperature observed from March to October.
Cumulative precipitation (Jun - July)	The accumulation of precipitation over June to July measured in mm.
Cooling Degree Days (CDD)(°F)	It represents how much warmer the mean air temperature is compared to a baseline temperature.

Table 2. Median model coefficients from the geographically weighted regression model with Land use/Land Cover (LULC) and climate variables as the explanatory variables. An extended summary statistic of the model coefficients is available in Table S2.

	Median coefficient
Intercept	1.64
All croplands fraction (%) in HUC12	3.03
Forest and shrubland fraction (%) in HUC8	-2.05
Grassland and pasture fraction (%) in HUC10	7.49
Wetland fraction (%) in HUC12	0.31
Cum. CDD (Mar-Oct) (°F)	-16.66
PDSI above normal (% area)	-1.21
T_{max} (May_Oct) (°C)	10.31
Cum. Precip (Jun-July) (Inch)	1.01
Residuals	-0.35
Local R^2	0.46

Funding Sources

NASA Ocean Biology and Biogeochemistry Program/Applied Sciences Program (proposals 14-SMDUNSOL14- 0001 and SMDSS20-0006)

Acknowledgement

This material is based upon work supported by the NASA Ocean Biology and Biogeochemistry Program/Applied Sciences Program (proposals 14-SMDUNSOL14- 0001 and SMDSS20-0006) and by the US EPA, NOAA, U.S. Geological Survey Toxic Substances Hydrology Program. We thank the three anonymous reviewers for providing constructive comments and suggestions, which helped significantly improve the manuscript's quality. This article has been reviewed by the Center for Environmental Measurement and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the US Government. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the US EPA. Authors declare no competing interests.

References

Ansko I, Kuusk J, Vendt R, Vabson V. MERIS Validation and Algorithm 4th reprocessing–MERIS Validation Team (MVT). Tartu Observatory, Tõravere, INTERMEDIATE REPORT 2015.

Armal S, Devineni N, Khanbilvardi R. Trends in extreme rainfall frequency in the contiguous United States: Attribution to climate change and climate variability modes. Journal of Climate 2018; 31: 369-385.

Baker D, Confesor R, Ewing D, Johnson L, Kramer J, Merryfield B. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: The importance of bioavailability. Journal of Great Lakes Research 2014; 40: 502-517.

Brooks BW, Lazorchak JM, Howard MD, Johnson MVV, Morton SL, Perkins DA, et al. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? Environmental toxicology and chemistry 2016; 35: 6-13.

Brunsdon C, Fotheringham A, Charlton M. Geographically weighted summary statistics—a framework for localised exploratory data analysis. Computers, Environment and Urban Systems 2002; 26: 501-524.

Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical analysis 1996; 28: 281-298.

Carmichael WW, Azevedo S, An JS, Molica R, Jochimsen EM, Lau S, et al. Human fatalities from cyanobacteria: chemical and biological evidence for cyanotoxins. Environmental health perspectives 2001; 109: 663-668.

Chen R-C, Dewi C, Huang S-W, Caraka RE. Selecting critical features for data classification based on machine learning methods. Journal of Big Data 2020; 7: 1-26.

Chorus I, Welker M. Exposure to cyanotoxins: Understanding it and short-term interventions to prevent it. Toxic Cyanobacteria in Water. CRC Press, 2021, pp. 295-400.

Clark JM, Schaeffer BA, Darling JA, Urquhart EA, Johnston JM, Ignatius AR, et al. Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. Ecological Indicators 2017; 80: 84-95.

Coffer MM, Schaeffer BA, Darling JA, Urquhart EA, Salls WB. Quantifying national and regional cyanobacterial occurrence in US lakes using satellite remote sensing. Ecological Indicators 2020; 111: 105976.

Coffer MM, Schaeffer BA, Salls WB, Urquhart E, Loftin KA, Stumpf RP, et al. Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales. Ecological indicators 2021; 128: 107822.

Cohen J. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates. Hillsdale, NJ 1988: 20-26.

CyAN. Cyanobacteria Assessment Network project page at NASA, Last accessed on 12/15/2022, https://oceancolor.gsfc.nasa.gov/projects/cyan/. 2022.

D'alba L, Colagrande P. MERIS Instrument Processing Facility Evolution, Technical Note. European Space Agency, Frascati, Italy, March, 2005.

ESA. Sentinel-3 User Guides. Last accessed on 06/01/2013. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-olci/overview/heritage. 2023.

Fotheringham AS, Charlton M, Brunsdon C. Two techniques for exploring non-stationarity in geographical data. Geographical Systems 1997; 4: 59-82.

Fotheringham AS, Charlton ME, Brunsdon C. Spatial variations in school performance: a local analysis using geographically weighted regression. Geographical and environmental Modelling 2001; 5: 43-66.

Gobler CJ. Climate change and harmful algal blooms: insights and perspective. Harmful algae 2020; 91: 101731.

Gregor J, Maršálek B. Freshwater phytoplankton quantification by chlorophyll a: a comparative study of in vitro, in vivo and in situ methods. Water research 2004; 38: 517-522.

Hallegraeff GM, Anderson DM, Belin C, Bottein M-YD, Bresnan E, Chinain M, et al. Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts. Communications Earth & Environment 2021; 2: 1-10.

Havens KE, Paerl HW. Climate change at a crossroad for control of harmful algal blooms. Environmental Science and Technology 2015; 49: 2.

Hirsch RM, Slack JR. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 1984; 20: 727-732.

Ho JC, Michalak AM, Pahlevan N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. Nature 2019; 574: 667-670.

Hou X, Feng L, Dai Y, Hu C, Gibson L, Tang J, et al. Global mapping reveals increase in lacustrine algal blooms over the past decade. Nature Geoscience 2022: 1-5.

HUC-USGS. Hydrologic Unit Maps. Last accessed on 06/01/2023, https://water.usgs.gov/GIS/huc.html. 2023.

Iiames J, Salls W, Mehaffey M, Nash M, Christensen J, Schaeffer B. Modeling Anthropogenic and Environmental Influences on Freshwater Harmful Algal Bloom Development Detected by MERIS Over the Central United States. Water Resources Research 2021; 57: e2020WR028946.

Kang T, Wang H, He Z, Liu Z, Ren Y, Zhao P. The effects of urban land use on energy-related CO2 emissions in China. Science of The Total Environment 2023; 870: 161873.

Karl TR, Koss WJ. Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983. Historical Climatology Series 4-3, National Climatic Data Center, Asheville, NC, 38 pp. National Oceanic and Atmospheric Administration (NOAA) 1984.

Kendall MG. A new measure of rank correlation. Biometrika 1938; 30: 81-93.

Kosten S, Huszar VL, Bécares E, Costa LS, van Donk E, Hansson LA, et al. Warmer climates boost cyanobacterial dominance in shallow lakes. Global Change Biology 2012; 18: 118-126.

Kudela R, Berdalet E, Bernard S, Burford M, Fernand L, Lu S, et al. Harmful algal blooms. A scientific summary for policy makers. IOC. UNESCO, IOC/INF-1320, 20pp 2015.

Kutser T. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. Limnology and Oceanography 2004; 49: 2179-2189.

Lam L, Suen S. Application of majority voting to pattern recognition: an analysis of its behavior and performance. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 1997; 27: 553-568.

Li Q, Sun W, Yun X, Huang B, Dong W, Wang XL, et al. An updated evaluation of the global mean land surface air temperature and surface temperature trends based on CLSAT and CMST. Climate Dynamics 2021; 56: 635-650.

Loftin KA, Graham JL, Hilborn ED, Lehmann SC, Meyer MT, Dietze JE, et al. Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. Harmful Algae 2016; 56: 77-90.

Lunetta RS, Schaeffer BA, Stumpf RP, Keith D, Jacobs SA, Murphy MS. Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. Remote Sensing of Environment 2015; 157: 24-34.

McKay L, Bondelid T, Dewald T, Johnston J, Moore R, Rea A. NHDPlus Version 2: User Guide. 2012.

Michalak AM, Anderson EJ, Beletsky D, Boland S, Bosch NS, Bridgeman TB, et al. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc Natl Acad Sci U S A 2013; 110: 6448-52.

Mishra S, Stumpf RP, Schaeffer B, Werdell PJ, Loftin KA, Meredith A. Evaluation of a satellite-based cyanobacteria bloom detection algorithm using field-measured microcystin data. Sci Total Environ 2021; 774: 145462.

Mishra S, Stumpf RP, Schaeffer BA, Werdell PJ, Loftin KA, Meredith A. Measurement of Cyanobacterial Bloom Magnitude using Satellite Remote Sensing. Scientific Reports 2019; 9: 1-17.

Myer MH, Urquhart E, Schaeffer BA, Johnston JM. Spatio-temporal modeling for forecasting high-risk freshwater cyanobacterial harmful algal blooms in Florida. Frontiers in environmental science 2020; 8: 581091.

NASS-USDA. Cropland Data Layer - National Download, Last accessed on 01/05/2022, https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php. 2022.

NCDC-NOAA. U.S. Climate Extremes Index (CEI), Last accessed on 01/02/2023, https://www.ncdc.noaa.gov/extremes/cei/. 2022.

NOAA-NCPC. Aggregated climate data within the U.S. limate divisions, Last accessed on 12/01/2011, https://www.ncei.noaa.gov/pub/data/cirs/climdiv/. 2022.

Paerl HW. Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. Limnology and Oceanography 1988; 33: 823-843.

Paerl HW, Huisman J. Blooms like it hot. Science 2008; 320: 57-58.

Qin B, Zhou J, Elser JJ, Gardner WS, Deng J, Brookes JD. Water depth underpins the relative roles and fates of nitrogen and phosphorus in lakes. Environmental science & technology 2020; 54: 3191-3198.

Sawilowsky SS. New effect size rules of thumb. Journal of modern applied statistical methods 2009; 8: 26.

Schaeffer BA, Loftin KA, Stumpf RP, Werdell PJ. Agencies collaborate, develop a cyanobacteria assessment network. EOS-Earth & Space Science News 2015; 96.

Schaeffer BA, Urquhart E, Coffer M, Salls W, Stumpf RP, Loftin KA, et al. Satellites quantify the spatial extent of cyanobacterial blooms across the United States at multiple scales. Ecological Indicators 2022; 140: 108990.

Seaber PR, Kapinos FP, Knapp GL. Hydrologic unit maps. Vol 2294: US Government Printing Office Washington, DC, USA, 1987.

Seegers BN, Werdell PJ, Vandermeulen RA, Salls W, Stumpf RP, Schaeffer BA, et al. Satellites for long-term monitoring of inland US lakes: The MERIS time series and application for chlorophyll-a. Remote Sensing of Environment 2021; 266: 112685.

Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American statistical association 1968; 63: 1379-1389.
Stroming S, Robertson M, Mabee B, Kuwayama Y, Schaeffer B. Quantifying the human health benefits of using satellite information to detect cyanobacterial harmful algal blooms and manage recreational advisories in US Lakes. GeoHealth 2020; 4: e2020GH000254.
Stumpf RP, Davis TW, Wynne TT, Graham JL, Loftin KA, Johengen TH, et al. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. Harmful Algae 2016a; 54: 160-173.
Stumpf RP, Johnson LT, Wynne TT, Baker DB. Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. Journal of Great Lakes Research 2016b; 42: 1174-1183.
Stumpf RP, Wynne TT, Baker DB, Fahnenstiel GL. Interannual variability of cyanobacterial blooms in Lake Erie. PLoS One 2012; 7: e42444.
Tomlinson MC, Stumpf RP, Wynne TT, Dupuy D, Burks R, Hendrickson J, et al. Relating chlorophyll from cyanobacteria-dominated inland waters to a MERIS bloom index. Remote Sensing Letters 2016; 7: 141-149.
Trees CC, Kennicutt II MC, Brooks JM. Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments. Marine Chemistry 1985; 17: 1-12.
Urquhart EA, Schaeffer BA. Envisat MERIS and Sentinel-3 OLCI satellite lake biophysical water quality flag dataset for the contiguous United States. Data in brief 2020; 28: 104826.
Vicent J, Sabater N, Tenjo C, Acarreta JR, Manzano M, Rivera JP, et al. FLEX end-to-end mission performance simulator. IEEE Transactions on Geoscience and Remote Sensing 2016; 54: 4215-4223.
Wells ML, Karlson B, Wulff A, Kudela R, Trick C, Asnaghi V, et al. Future HAB science: Directions and challenges in a changing climate. Harmful Algae 2020; 91: 101632.
Whitman P, Schaeffer B, Salls W, Coffer M, Mishra S, Seegers B, et al. A validation of satellite derived cyanobacteria detections with state reported events and recreation advisories across US lakes. Harmful Algae 2022; 115: 102191.
Wilkinson GM, Walter JA, Buelo CD, Pace ML. No evidence of widespread algal bloom intensification in hundreds of lakes. Frontiers in Ecology and the Environment 2022; 20: 16-21.
Wynne T, Meredith A, Briggs T, Litaker W, Stumpf R. Harmful Algal Bloom Forecasting Branch Ocean Color Satellite Imagery Processing Guidelines. NOAA Technical Memorandum NOS NCCOS 2018; 252: 48.
Wynne TT, Mishra S, Meredith A, Litaker RW, Stumpf RP. Intercalibration of MERIS, MODIS, and OLCI Satellite Imagers for Construction of Past, Present, and Future Cyanobacterial Biomass Time Series. Remote Sensing 2021; 13: 2305.
Wynne TT, Stumpf RP. Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in western lake erie from 2002–2014. Toxins 2015; 7: 1649-1663.
Wynne TT, Stumpf RP, Tomlinson MC, Dyble J. Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data. Limnology and Oceanography 2010; 55: 2025-2036.
Wynne TT, Stumpf RP, Tomlinson MC, Warner RA, Tester PA, Dyble J, et al. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. International Journal of Remote Sensing 2008; 29: 3665-3672.

Zhang J, Phaneuf DJ, Schaeffer BA. Property values and cyanobacterial algal blooms: Evidence from satellite monitoring of Inland Lakes. Ecological Economics 2022; 199: 107481.

Zurita-Milla R, Clevers J, Schaepman ME, Kneubuehler M. Effects of MERIS L1b radiometric calibration on regional land cover mapping and land products. International Journal of Remote Sensing 2007; 28: 653-673.