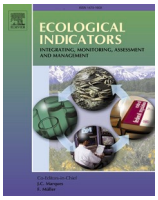
Ecological Indicators 128 (2021) 107822

Contents lists available at ScienceDirect

Ecological Indicators

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



Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales

Megan M. Coffer [a,b], Blake A. Schaeffer [c], Wilson B. Salls [c], Erin Urquhart [d], Keith A. Loftin [e], Richard P. Stumpf [f], P. Jeremy Werdell [g], John A. Darling [c]

[a] ORISE Fellow, U.S. EPA, Office of Research and Development, Durham, NC, USA
[b] Center for Geospatial Analytics, North Carolina State University, Raleigh, NC, USA
[c] U.S. EPA, Office of Research and Development, Durham, NC, USA
[d] Science Systems and Applications, Inc., Ocean Ecology Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[e] U.S. Geological Survey, Kansas Water Science Center, Lawrence, KS, USA
[f] National Oceanic and Atmospheric Administration, National Centers for Coastal Ocean Science, Silver Spring, MD, USA
[g] Ocean Ecology Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

ARTICLE INFO

Keywords:
Satellite remote sensing
Cyanobacteria
Inland waters
Water quality
Clustering analysis
Indicator

ABSTRACT

Cyanobacterial blooms can have negative effects on human health and local ecosystems. Field monitoring of cyanobacterial blooms can be costly, but satellite remote sensing has shown utility for more efficient spatial and temporal monitoring across the United States. Here, satellite imagery was used to assess the annual frequency of surface cyanobacterial blooms, defined for each satellite pixel as the percentage of images for that pixel throughout the year exhibiting detectable cyanobacteria. Cyanobacterial frequency was assessed across 2,196 large lakes in 46 states across the continental United States (CONUS) using imagery from the European Space Agency's Ocean and Land Colour Instrument for the years 2017 through 2019. In 2019, across all satellite pixels considered, annual bloom frequency had a median value of 4% and a maximum value of 100%, the latter indicating that for those satellite pixels, a cyanobacterial bloom was detected by the satellite sensor for every satellite image considered. In addition to annual pixel-scale cyanobacterial frequency, results were summarized at the lake- and state-scales by averaging annual pixel-scale results across each lake and state. For 2019, average annual lake-scale frequencies also had a maximum value of 100%, and Oregon and Ohio had the highest average annual state-scale frequencies at 65% and 52%. Pixel-scale frequency results can assist in identifying portions of a lake that are more prone to cyanobacterial blooms, while lake- and state-scale frequency results can assist in the prioritization of sampling resources and mitigation efforts. Satellite imagery is limited by the presence of snow and ice, as imagery collected in these conditions are quality flagged and discarded. Thus, annual bloom frequencies within nine climate regions were investigated to determine whether missing data biased results in climate regions more prone to snow and ice, given that their annual summaries would be weighted toward the summer months when cyanobacterial blooms tend to occur. Results were unbiased by the time period selected in most climate regions, but a large bias was observed for the Northwest Rockies and Plains climate region. Moderate biases were observed for the Ohio Valley and the Southeast climate regions. Finally, a clustering analysis was used to identify areas of high and low cyanobacterial frequency across CONUS based on average annual lake-scale cyanobacterial frequencies for 2019. Several clusters were identified that transcended state, watershed, and eco-regional boundaries. Combined with additional data, results from the clustering analysis may offer insight regarding large-scale drivers of cyanobacterial blooms.

1. Introduction

Cyanobacterial blooms can have negative effects on human health and local ecosystems. Human exposure to cyanobacteria can occur through either recreational exposure or the consumption of contaminated drinking water, and can cause skin and eye irritation, damage to liver and kidney function, and flu-like symptoms (World Health Organization, 2003). Effects of cyanobacteria and their toxins on ecosystems have been documented for bacteria, algae, plants, fish, and aquatic invertebrates (Zanchett and Oliveira-Filho, 2013). In the United States, reports of human or animal illness associated with cyanobacterial blooms have occurred in 43 of 50 states (Dubrovsky et al., 2016).

Economic impacts of cyanobacteria result in part from water treatment costs, medical treatment costs, and surveillance and monitoring

https://doi.org/10.1016/j.ecolind.2021.107822
Received 29 January 2021; Received in revised form 19 May 2021; Accepted 22 May 2021
Available online 26 May 2021

costs. As an example, treatment costs for toxin removal vary, but totals of $6,000 to $7,000 per day to treat drinking water supplying a city the size of Toledo, Ohio (population of approximately 280,000), were reported in 2013 (Great Lakes Coastal Resilience, 2013). Additional costs can be incurred through the treatment of taste-and-odor compounds and disinfection byproducts. Costs associated with field monitoring are more difficult to quantify, but one estimate from the Washington State Department of Ecology, which manages and budgets a state program for harmful algal bloom monitoring, estimated laboratory costs of event response samples at $60,000 per year and the staff time to facilitate and coordinate the program at $150,000 per year (Anderson-Abbs et al., 2016).

Satellite remote sensing imagery has shown utility as a spatially and temporally efficient monitoring technique for inland cyanobacterial blooms in the United States, at both the regional scale (Clark et al., 2017; Urquhart et al., 2017) and the national scale (Coffer et al., 2020). Satellite imagery cannot be used to detect toxins (Stumpf et al., 2016) nor can it be used to characterize water quality at depth, as the algorithms typically used for cyanobacteria can detect blooms up to a depth of 2 m in clear water (Mishra et al., 2005) and much less than 2 m in more turbid waters (Wynne et al., 2010). However, satellites can be used to effectively evaluate surface cyanobacterial presence above the satellite's minimum reporting level (Kutser, 2009). Satellite imagery can provide consistent, near real-time assessments of water quality specific to cyanobacterial blooms, which can alert water quality managers to potential concerns and emerging risks (Schaeffer et al., 2018). Moreover, satellite imagery can characterize the full population of phytoplankton present in a waterbody compared to traditional phytoplankton microscopic examination, which does not routinely include picoplankton because of their small size (Śliwińska-Wilczewska et al., 2018). As a result of expanded coverage, the availability of satellite imagery has been associated with cost savings specific to cyanobacterial monitoring (Papenfus et al., 2020). Using a 2017 cyanobacterial bloom in Utah Lake, Utah, as a case study, Stroming et al. (2020) estimated that the use of satellite data for harmful algal bloom monitoring resulted in approximately $370,000 in cost savings related to human health outcomes in that single event.

This study uses satellite imagery to build on regional results presented in Clark et al. (2017) by assessing annual cyanobacterial frequency at multiple spatial scales. At the smallest spatial scale considered, annual pixel-scale cyanobacterial frequency can be computed as the percentage of satellite observations throughout the year that indicate a cyanobacterial bloom is present. At broader spatial scales, these pixel-scale results can be summarized into average annual lake-scale and state-scale cyanobacterial frequencies by averaging annual results for all pixels contained in each lake and state. Previous studies have used satellite imagery to present annual cyanobacterial bloom frequencies at various regional scales: Clark et al. (2017) analyzed bloom frequency for lakes in the states of Ohio and Florida; Kahru et al. (2007) assessed bloom frequency in the Baltic Sea; both Hu et al. (2010) and Yang et al. (2016) quantified the temporal frequency of cyanobacterial blooms for Taihu Lake in China. Despite the use of a cyanobacterial frequency metric in past studies, the effect of missing observations on annual summaries has not been documented. While satellite imagery has the potential to provide frequent, year-round observations, cold-season data can be limited as observations are often quality flagged and discarded due to snow and ice cover (Coffer et al., 2020). Quantifying this effect can demonstrate if annual frequency metrics are appropriate at all lakes and if comparisons across lakes or states in differing regions are valid.

Here, surface cyanobacterial bloom frequency is presented at the pixel-, lake-, and state-scales using satellite observations from over 2,000 lakes across the contiguous United States (CONUS). State boundaries were selected as opposed to watershed boundaries to align with typical state management approaches. Average annual lake-scale results were used to explore two additional components of the dataset. First, potential biases caused by the inclusion of cold-season data in annual summaries were explored. This was done by testing the hypothesis that summary metrics would be biased toward summertime data, when cyanobacterial blooms tend to occur, resulting in an artificially higher annual cyanobacterial bloom frequency metric in areas affected by snow and ice. Second, spatial patterns were explored to identify spatial clusters of high and low bloom frequency lakes. This information may be useful for prioritizing management efforts as well as for understanding large-scale drivers of cyanobacterial blooms. The analysis of drivers was beyond the scope of this effort but has been reported through other analyses at the regional level (Myer et al., 2020). This study addresses the following research objectives:

1. Characterize the temporal frequency of surface cyanobacterial blooms across CONUS at the satellite pixel-, lake-, and state-scales for the years 2017 through 2019.
2. Explore potential biases caused by the inclusion of cold-season data. This was done by comparing average annual lake-scale results to lake-scale results computed using a temporal subset to represent the snow-free period.
3. Identify spatial clusters of lakes with high cyanobacterial frequency and lakes with low cyanobacterial frequency using average annual lake-scale results.

2. Data and methods

2.1. Satellite observations

Satellite observations were obtained from the European Space Agency's (ESA's) Ocean and Land Colour Instrument (OLCI) onboard the Sentinel-3A satellite. Sentinel-3A was launched in February 2016. Continuous observations spanning the full year over CONUS became available beginning in 2017 and are expected to be available for similar analyses in future years. OLCI offers a 2 to 3 day revisit period, 12-bit radiometric resolution, and 300-m spatial resolution at nadir, where nadir is defined as the point on Earth's surface directly below the satellite. Imagery is collected in 21 spectral bands spanning 400 nm to 1020 nm.

OLCI spectral surface reflectance imagery was obtained from the National Aeronautics and Space Administration (NASA) Ocean Biology Processing Group (OBPG; https://oceandata.sci.gsfc.nasa.gov) and was further processed and quality assured as described in Urquhart and Schaeffer (2020). The OBPG redistributes this data record through a data sharing agreement between NASA and ESA. Briefly, the OBPG acquired Level-1B (calibrated, top-of-atmosphere) data directly from ESA and processed them to Level-2 imagery through removal of the contribution of spectral Rayleigh scattering from the top-of-atmosphere signal using NASA's standard data processing software packaged as part of the SeaWIFS Data Analysis System (SeaDAS; https://seadas.gsfc.nasa.gov). A high resolution (approximately 60-m) land mask based on the NASA Shuttle Radar Topography Mission Water Body Data Shapefiles (NASA JPL, 2013) was used, with modifications by Urquhart and Schaeffer (2020) to correct for embedded inaccuracies in the dataset, such as missing lakes and reservoirs in Rhode Island and Massachusetts. Several processing masks were applied to exclude questionable Level-2 data. A cloud flag was adopted following Wynne et al. (2018) and a flag for mixed land and water pixels was developed to identify cases where the land mask reported a water pixel, but that pixel did not contain water at the time of satellite observation. Flags to indicate potential contamination due to adjacency effects and to identify snow- or ice-covered waterbodies were also applied (Wynne et al., 2018). In summary, satellite pixels containing mixed land and water, Sun glint, cloud cover, or those adjacent to a bright target, such as land, snow, or ice, were discarded. Satellite observations were aggregated to weekly composites that preserved the maximum data value for each pixel. Hereafter, a valid satellite pixel refers to a satellite pixel in a weekly composite that was not discarded due to any of these exclusion criteria.

Cyanobacterial abundance was characterized for each valid satellite pixel using the CI-cyano algorithm (Lunetta et al., 2015; Wynne et al., 2010, 2008). Spectral bands centered at 665 nm, 681 nm, and 709 nm are used to assess bloom biomass and those centered at 620 nm, 665 nm, and 681 nm are used as exclusion criteria to prevent the quantification of non-cyanobacterial blooms. The progression of the CI-cyano algorithm is detailed in Coffer et al. (2020). Throughout this study, a pixel is classified as a cyanobacterial detection if the CI-cyano algorithm returns any detectable value, indicating cyanobacteria in concentrations above the detection limit of the sensor. As a final step, all satellite retrievals were remapped onto a consistent Albers conic projection with a 300-m bin size to ensure spatial consistencies across all analyses.

CI-cyano has demonstrated success at the national (Coffer et al., 2020) and state (Clark et al., 2017; Mishra et al., 2019; Urquhart et al., 2017) scales for detecting cyanobacteria. Quantitative validation of CI-cyano has been assessed against cyanobacterial cell counts and concentrations of chlorophyll-a across 38 CONUS states (Clark et al., 2017; Lunetta et al., 2015). Additionally, qualitative validation of CI-cyano has been assessed against cyanobacteria presence and absence (Mishra et al., 2021; Schaeffer et al., 2018). Using CI-cyano, Coffer et al. (2020) demonstrated expected seasonality across 46 CONUS states.

Given the spatial resolution of OLCI, a total of 2,196 lakes and reservoirs across 46 states can be resolved (Urquhart and Schaeffer, 2020). Hereafter, these are collectively referred to as lakes. For a lake to be considered resolvable, it must be of sufficient size and shape to accommodate at least one 300-m water-only satellite pixel after applying the exclusion criteria described above. These lakes range in surface area from 1.3 square kilometers (km^2) to over 4,000 km^2, limiting this analysis to relatively large lakes. West Virginia and Delaware are the only states in CONUS that do not contain any resolvable lakes. The number of valid satellite pixels in these resolvable lakes had a median value of 19 and ranged from 1 satellite pixel to nearly 40,000 satellite pixels (Figure S1).

2.2. Computing cyanobacterial frequency

The proportion of valid pixels that returned detectable cyanobacteria was used to determine annual cyanobacterial frequencies at different scales (Fig. 1). Cyanobacterial bloom frequency was computed following Clark et al. (2017) for each satellite pixel as the percentage of weekly satellite composites exhibiting cyanobacterial presence above the minimum detection limit of the satellite sensor relative to the total number of weekly satellite composites that contained a valid measurement (i.e., measurements that were not quality flagged and discarded) for the given pixel (Fig. 1A; Eq. (1)):

$$\text{Annual pixel-scale frequency} = 100\% \times \frac{\text{n of pixels with detectable CI-cyano}}{\text{n of all valid pixels}} \quad (1)$$

In addition to pixel-scale bloom frequencies, average annual lake-scale bloom frequencies and average annual state-scale bloom frequencies were computed. For each year of observations, average annual lake-scale bloom frequencies were computed by averaging annual pixel-scale bloom frequencies for all pixels contained within a lake (Fig. 1B; Eq. (2)). Pixels within a smaller surface area lake carry a larger weight than pixels within a larger surface area lake when computing lake-scale averages. Average annual state-scale bloom frequencies were computed by averaging annual pixel-scale bloom frequencies for all pixels contained within a state (Fig. 1C; Eq. (3)). Lakes with a larger surface area carry a larger weight than lakes with a smaller surface area when computing state-scale averages.

$$\text{Annual lake-scale frequency} = \frac{\sum \text{annual pixel-scale frequencies within lake}}{\text{n of pixels within lake}} \quad (2)$$

$$\text{Annual state-scale frequency} = \frac{\sum \text{annual pixel-scale frequencies within state}}{\text{n of pixels within state}} \quad (3)$$

Annual pixel-scale cyanobacterial frequencies were computed using a range of denominators because each denominator reflected the number of weekly composites that contained a valid observation for the given satellite pixel. Denominators ranged from 1 to 52 representing the 52 weeks in a calendar year (Figure S2). Pixel-scale cyanobacterial frequencies computed with a small sample size (i.e., a small denominator) may not necessarily be representative of conditions across the entire year. Moreover, if pixel-scale frequencies across a lake or state were computed using different sample sizes, average lake- and state-scale cyanobacterial frequencies could be aggregating different time periods in their annual summaries.

The focus of this paper is cyanobacterial frequency computed by

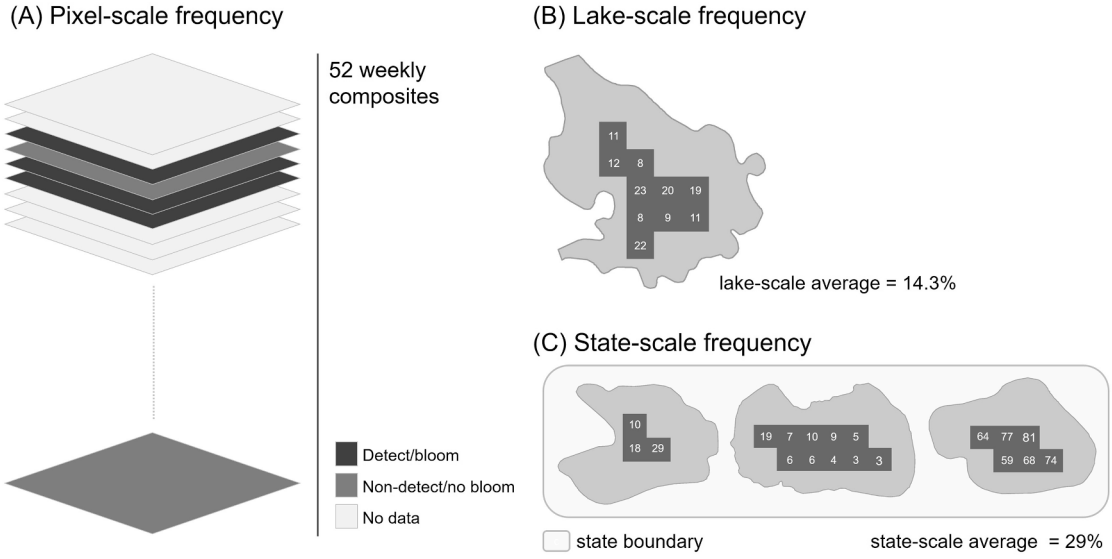


Fig. 1. Conceptual diagrams illustrating the computation of (A) annual pixel-scale frequency, (B) average annual lake-scale frequency, and (C) average annual state-scale frequency. Example pixel-scale frequency values and the resulting average annual lake- and state-scale frequencies are shown for (B) and (C), respectively.

considering cyanobacterial presence and absence, but supplemental information also includes cyanobacterial frequency within each of the risk levels defined by the World Health Organization (WHO). The WHO states that cyanobacteria concentrations of up to 20,000 cells/mL correspond to a relatively low probability of adverse health effects in recreational waters, those of up to 100,000 cells/mL correspond to a moderate probability, and those above 100,000 cells/mL correspond to a high probability (World Health Organization, 2003).

It should be noted that pixel-scale estimates primarily characterize the center of the lake, and narrow reaches are unobservable using 300-m satellite imagery. The relatively coarse spatial resolution of OLCI results in the loss of smaller lakes as well as more narrow portions of resolvable lakes. Lake-scale summaries exclude satellite pixels along the land-water interface and exclude a one-pixel buffer from the shoreline, an area where blooms tend to accumulate, particularly in windy conditions (Gons et al., 2005). State-scale summaries consider only lakes of sufficient size and shape to accommodate the spatial resolution of OLCI. Thus, average state-scale bloom frequencies presented here could be biased compared to state-scale analyses that include a more complete representation of lakes throughout the state.

2.3. Analyzing seasonal effects on cyanobacterial frequency

Missing data occurs primarily due to cloud cover, Sun glint, or snow and ice. Across CONUS, northern latitude states are more affected by snow and ice cover, particularly in the winter months. Eliminating weekly composites containing snow and ice decreases the number of weekly composites considered in the annual bloom frequency calculation, thus potentially artificially inflating the resulting frequency value by weighting results more toward summer observations, a time when cyanobacterial blooms tend to occur. Thus, annual bloom frequency computed using weekly composites for each week in the year was compared to bloom frequency computed using a temporal subset of weekly composites representing only periods with no more than 10% snow and ice cover across CONUS. This analysis indicates whether annual bloom frequency was artificially inflated by the inclusion of the cold-season period due to a bias toward summer observations.

Relatively snow-free weekly composites were defined as those in which no more than 10% of the area of CONUS was covered in snow or ice. Shapefiles indicating daily snow and ice cover at a 4-km resolution were obtained from the National Snow and Ice Data Center (National Ice Center, 2008). This dataset provides maps of snow cover and sea ice for the Northern Hemisphere from February 1997 to the present and are derived from a variety of data products, including satellite imagery and *in situ* data. These shapefiles were formatted according to Urquhart and Schaeffer (2020): holes in the original shapefiles were filled before creating weekly composites preserving the maximum snow and ice extent for each week and rasterizing the shapefiles to match the spatial characteristics of OLCI imagery. Bloom frequency across CONUS was then computed as described in Section 2.2 for the subset of weekly composites. While these weekly composites do not represent a fully snow-free period, they will hereafter be referred to as the snow-free period.

Cyanobacterial frequency computed using the snow-free period could have a different result than cyanobacterial frequency computed using the entire annual time period. Therefore, the two populations were compared by looking at the differences between each pair of annual and snow-free results. The Wilcoxon signed-rank test is a nonparametric approach to test whether the corresponding data population distributions are identical, using each pair of observations and without assuming they follow a normal distribution (Wilcoxon, 1945). The Wilcoxon signed-rank test was applied to lake-scale results within each of the nine climate regions defined by the National Center for Environmental Information (Figure S3; Karl and Koss, 1984) using cyanobacterial frequency results from 2019 and the *rstatix* package in R Version 4.0.0 (Kassambara, 2020; R Core Team, 2020). These nine climate regions group climatically consistent states within CONUS and were used here to group together states more prone to missing data due to snow and ice cover in the winter months. Lake centroids were used to assign each lake to a climate region.

Cohen's d was used to quantify the difference between the two datasets in terms of effect size (Cohen, 1992). Cohen's d was computed as the z-score divided by the square root of the sample size, where the z-score is the measure of how many standard deviations a raw score is below or above the population mean. Generally, an absolute value of Cohen's d between 0.1 and 0.3 indicates a small effect, an absolute value between 0.3 and 0.5 indicates a moderate effect, and an absolute value above 0.5 indicates a large effect.

2.4. Assessing spatial patterns in cyanobacterial frequency

Spatial clusters of lakes with high cyanobacterial bloom frequencies and spatial clusters of lakes with low cyanobacterial bloom frequencies were identified across CONUS. This analysis may offer important information for management prioritization and can assist in understanding large-scale drivers of cyanobacterial blooms. Clusters of high and low cyanobacterial bloom frequencies were identified via the Getis-Ord Gi* statistic (Getis and Ord, 1992; Ord and Getis, 1995). A lake must itself have a high average annual bloom frequency and also be surrounded by other lakes with high average annual bloom frequencies to be considered part of a cluster of high bloom frequencies. If a local group of lakes has a higher Getis-Ord Gi* statistic than the global average, it is considered a cluster of lakes with a high cyanobacterial bloom frequency. The same reasoning is used to identify clusters of lakes with low cyanobacterial bloom frequencies. The Getis-Ord Gi* statistic does not quantify the number of clusters present nor does it specify which cluster a given lake falls within, but clusters can be visually discerned from the statistical results. This statistic was computed based on 2019 average annual lake-scale bloom frequencies using the *Hot Spot Analysis (Getis-Ord Gi*)* tool in ArcGIS 10.7.1 (ESRI, 2019).

Input parameters for the Getis-Ord Gi* statistic were optimized for the 2019 average annual lake-scale bloom frequency dataset. First, the *Incremental Spatial Autocorrelation* tool was used to measure spatial autocorrelation. The first distance with a declining z-score was selected as the neighborhood search value (Figure S4). This neighborhood search value was then used to create a spatial weights matrix using the *Generate Spatial Weights Matrix* tool; the number of nearest neighbors was set to the recommended value of 8 and row standardization was applied which standardizes each spatial weight by the row sum, an approach recommended when feature distribution is potentially biased. Spatial weights from the resulting spatial weights matrix were used in the *Hot Spot Analysis (Getis-Ord Gi*)* tool to conceptualize the spatial relationships. Given a large sample size of 2,196 data points, the false discovery rate correction was applied, which reduces resulting p-values to account for multiple testing and spatial dependence. Each Getis-Ord Gi* statistic has an associated z-score and p-value. A large, positive z-score and a small p-value for a feature indicate a spatial clustering of high values. A large, negative z-score and a small p-value indicate a spatial clustering of low values.

3. Results and discussion

3.1. Annual pixel-scale cyanobacterial frequency

Pixel-scale surface bloom frequencies ranged from 0% to 100% across all pixels for the years 2017, 2018, and 2019 (Fig. 2). The median of the distribution for 2017 was 2.7%, which was lower compared to 2018 (median of 4.6%) and 2019 (median of 4.4%). These medians characterize, for all pixels, the percentage of valid weekly composites whose surface cyanobacterial abundance was above the detection limit of the sensor. A small (1% to 2%) proportion of pixels reached a frequency value of 100%, indicating a detectable cyanobacterial bloom

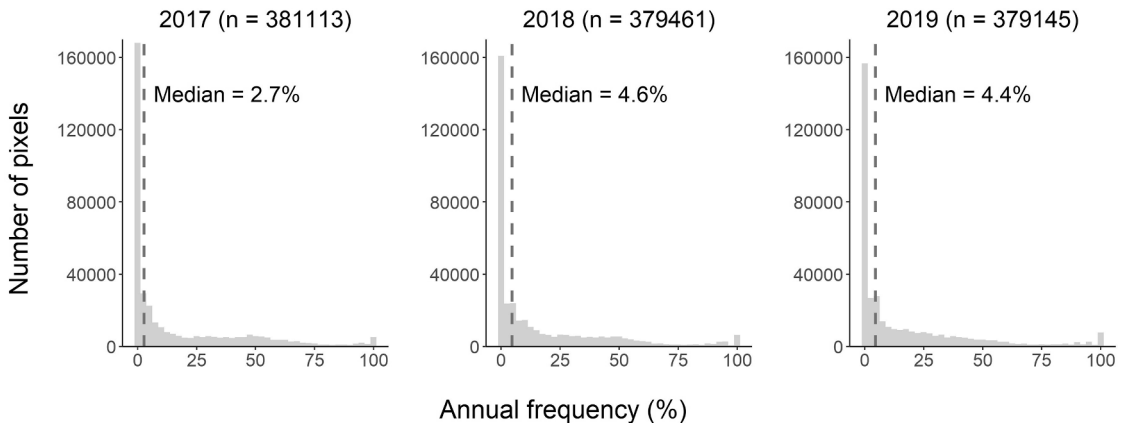


Fig. 2. Histogram of annual bloom frequency values for all satellite pixels across CONUS for the years 2017, 2018, and 2019. The median bloom frequency values for each year were 2.7%, 4.6%, and 4.4%, respectively.

was present for every valid weekly composite considered.

Cyanobacterial frequency was fairly stable across years, as each histogram exhibits a similar distribution. All years tended toward lower frequencies, with half of the distribution falling below a bloom frequency between approximately 3% and 5%. Generally, the distribution then continued to decline, reaching the lowest number of instances between bloom frequencies of 75% and bloom frequencies near 100%. A small peak is apparent for bloom frequencies of 100%, particularly for the years 2018 and 2019, which also exhibited the highest median values.

Annual pixel-scale frequencies can be used to assess relatively small-scale (300 m) variations in surface bloom dynamics, such as for a subset of lakes in North Dakota for the year 2019 (Fig. 3). This region of North Dakota was chosen given confirmation of bloom events for the year 2019 in Antelope Lake, Devils Lake, and Stump Lake (ND DEQ, 2019). Annual pixel-scale frequencies may assist in resource management and help inform the prioritization of sampling efforts at the lake-scale. At Dry Lake, for example, pixel-scale results indicated that the northern portion of the lake is more prone to bloom events, while the southern segment of the lake is characterized by lower frequencies. Across other lakes in North Dakota, pixel-scale annual frequencies ranged from low values across Cranberry Lake to high values across Lake Alice and Rush Lake. Devils Lake, the largest natural lake in the state, had fairly consistent cyanobacterial blooms spatially, with pixel-scale frequencies nearly all around 50%.

Clark et al. (2017) assessed pixel-scale bloom frequencies in Florida and Ohio using the same methods presented here, but for the entire 2008 through 2011 period collectively. Additionally, a threshold of 100,000 cells/mL was used as a demonstration to correspond to the WHO high-risk value rather than a threshold reflecting the detection limit of the sensor, which is presented here. In Florida, Clark et al. (2017) found pixel-scale bloom frequencies to average 30% for 2008 through 2011 whereas in Ohio, Clark et al. (2017) found the average to be 5%. That study also found the highest pixel-scale frequency observed for the state of Florida to be 99%, which occurred in Lake Apopka, and the highest pixel-scale frequency observed for the state of Ohio to be 83%, which occurred in Grand Lake (commonly referred to as Grand Lake St. Marys). The fact that these values are all lower than comparable ones reported here is consistent with expectations given the more restrictive bloom threshold used in Clark et al. (2017).

3.2. *Average annual lake-scale cyanobacterial frequency*

Average annual lake-scale frequencies varied drastically across CONUS for the year 2019 (Fig. 4). Over 68% of all resolvable lakes had an average annual bloom frequency falling in the first quartile (ranging between a frequency of 0% and 25%). The state of Maine had a particularly large number of resolvable lakes falling in this range. Across CONUS, approximately 20% of resolvable lakes had an average annual bloom frequency falling in the second quartile and 9% of resolvable

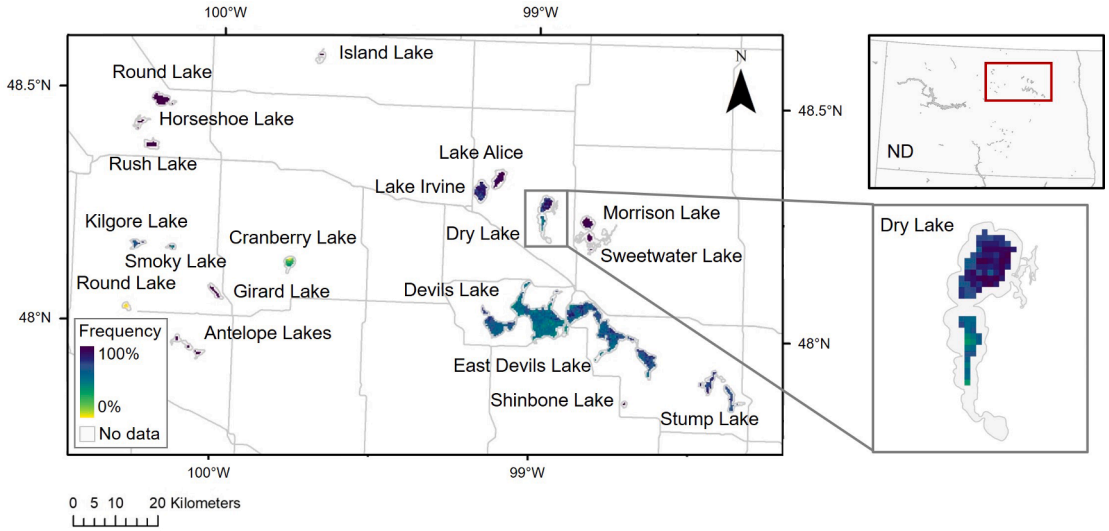


Fig. 3. Annual pixel-scale cyanobacterial bloom frequency for the year 2019 at a subset of lakes in North Dakota (ND).

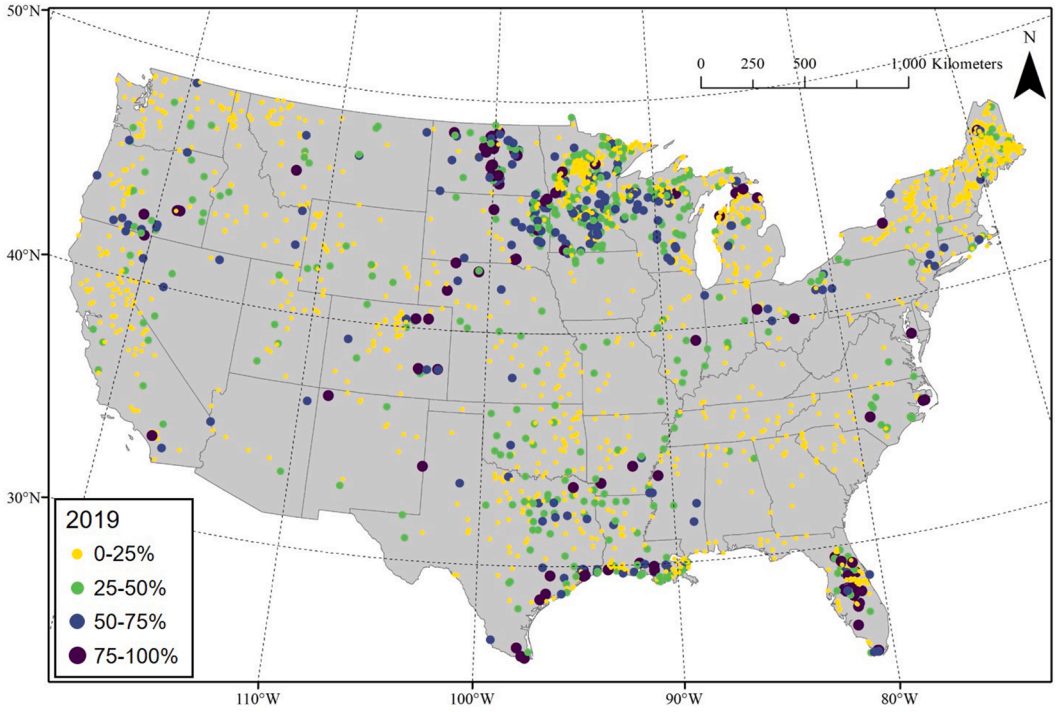


Fig. 4. Average annual lake-scale cyanobacterial bloom frequencies for the year 2019. Each point represents a lake centroid.

lakes fell in the third quartile. Just under 5% of resolvable lakes had an average annual bloom frequency exceeding 75%, with most of these lakes falling in Florida, the Texas Gulf Coast, and the Dakotas.

The National Lakes Assessment (NLA) uses field monitoring to assess the conditions of lakes, ponds, and reservoirs across the United States. The 2012 NLA included sampling for cyanobacteria and microcystins at over 1,000 lakes selected via stratified random sampling (U.S. EPA, 2016). Lakes were characterized as either least, moderately, or most disturbed based on deviations from the distribution of all studied lakes contained within a region. That study categorized 23% of studied lakes as moderately disturbed and 15% of studied lakes as most disturbed. The remaining 61% of studied lakes were categorized as least disturbed. While the categorization used by the 2012 NLA differed from the bloom frequency metric presented here, results from the two approaches are similar. Here, 68% of lakes fell in the lowest quartile of cyanobacterial bloom frequency compared to the 61% of NLA lakes categorized as lea in the st disturbed.

Average annual lake-scale frequencies for the years 2017 through 2019 indicated stability within this time period, with bloom frequencies between 0% and 25% for most lakes across CONUS in 2017 and 2018 (Figures S5-S6). A notable difference, however, occurred in lakes throughout the central United States and those along the Mississippi River. For the years 2017 and 2018, many of these lakes had higher average annual lake-scale bloom frequencies than for the year 2019. The National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center (NCDC) produces annual climate reports to analyze national temperature and precipitation and place the data into a historical perspective (NCDC, 2019). In 2019, several states in the Upper Midwest and Northwest Rockies and Plains climate regions experienced their wettest years on record, which included 125 years of observations as of 2019. This heavy precipitation upstream led to record flooding, resulting in damages of $20 billion across the Mississippi, Arkansas, and Missouri Rivers (NOAA NCEI, 2021). Consequently, lakes can become diluted from a large influx of freshwater which can reduce cyanobacterial blooms (Berger et al., 2008). Increased precipitation can also reduce water stratification which can lead to underrepresentation of cyanobacteria as the bloom becomes vertically distributed within the water column (Reynolds, 2006) while the satellite measures only the upper layer of the water column.

Presenting the distribution of average annual lake-scale bloom frequencies can assist in the identification of lakes within a state that are more prone to cyanobacterial blooms, as shown in Fig. 5 for a single state in each of the nine climate regions defined by Karl and Koss (1984): Oregon (OR) in the Northwest, Vermont (VT) in the Northeast, Nebraska (NE) in the Northwest Rockies and Plains, Utah (UT) in the Southwest, Iowa (IA) in the Upper Midwest, Texas (TX) in the South, Florida (FL) in the Southeast, Ohio (OH) in the Ohio Valley, and California (CA) in the West. For each state, several lakes and their associated cyanobacterial frequencies are labeled. These labeled lakes were chosen as a demonstration of the frequency metric as they have water quality information, including cyanobacterial detections, for either 2019 or previous years.

Several of the lakes labeled in Fig. 5 had relatively low average annual lake-scale cyanobacterial frequency for 2019, including Detroit Lake in Oregon, Lake Meredith and Grapevine Lake in Texas, and Lake Tahoe in California. Detroit Lake, which experienced a cyanobacterial bloom in 2018 upstream from a drinking water intake leading to the issuance of a “Do Not Drink” advisory (The Novak Consulting Group, 2018), had an annual frequency of 6%. Field observations support the presence of cyanobacteria in Lake Meredith and Grapevine Lake (Trevino and Petersen, 2020). Lake Tahoe, a highly studied ultra-oligotrophic lake in California known for its water clarity (e.g., Beutel and Horne, 2018; Coats et al., 2016), had no detectable cyanobacteria in any resolvable weekly composites across the lake as demonstrated through an average annual lake-scale bloom frequency of 0%.

Lakes labeled in Fig. 5 with high average annual bloom frequencies included Lakes Jesup and Apopka in Florida, Grand Lake in Ohio, and Lake Elsinore in California. This study found Lake Apopka and Grand Lake to both average the highest potential bloom frequency value of 100%. Using the same bloom frequency metric presented here, Clark et al. (2017) found bloom frequency in Lake Apopka to be the highest in the state of Florida, reaching a maximum value of 99% from 2008 through 2011. Similarly, Mishra et al. (2019) found Lake Apopka to be the highest in cyanobacterial bloom magnitude for the state of Florida from 2003 through 2011. Grand Lake is a heavily studied reservoir in

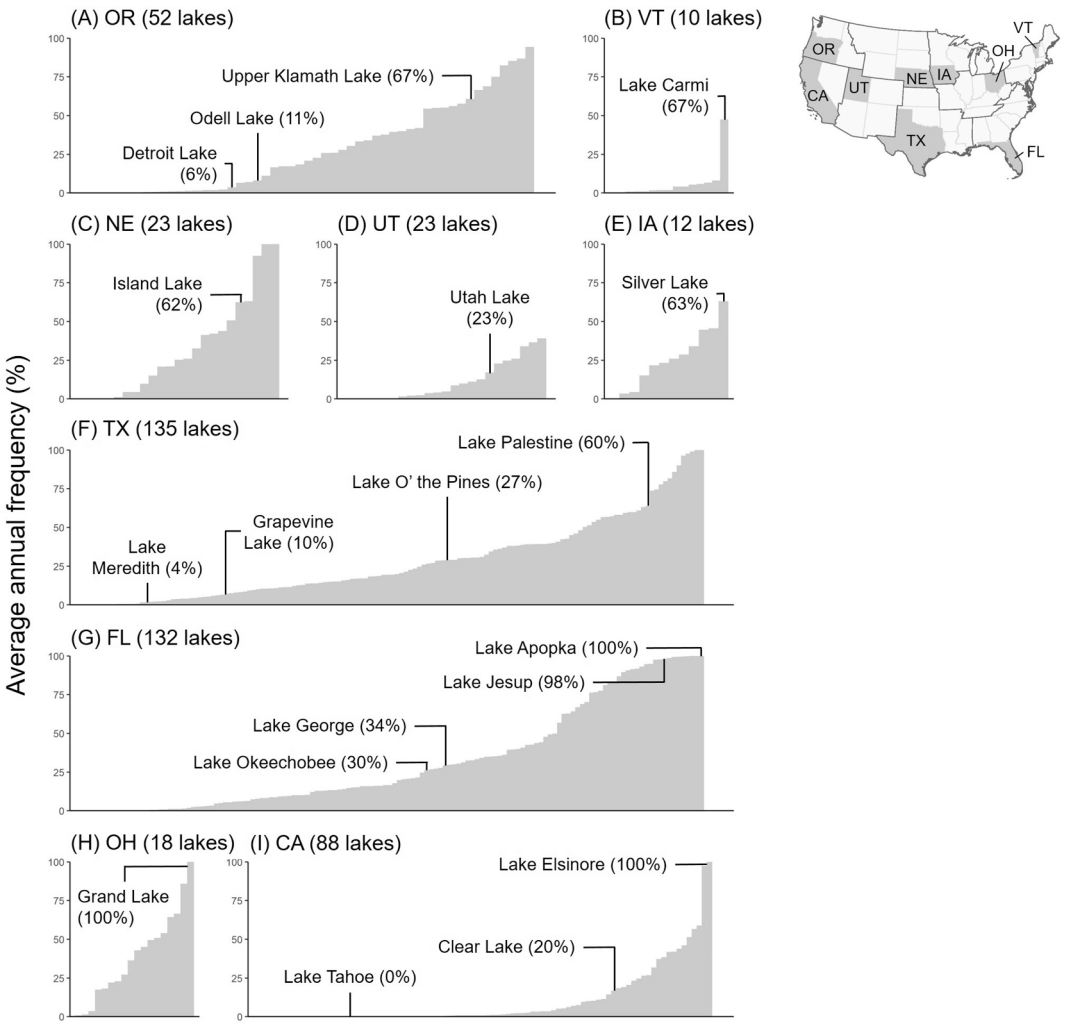


Fig. 5. Ranked distribution of average annual lake-scale cyanobacterial frequencies for the year 2019 for resolvable lakes in the states of (A) Oregon, (B) Vermont, (C) Nebraska, (D) Utah, (E) Iowa, (F) Texas, (G) Florida, (H) Ohio, and (I) California. For each state, some lakes with documented cases of cyanobacterial blooms either for 2019 or in previous years are labeled. The x-axis is unitless with a range from 1 to the total number of observable lakes in the given state.

western Ohio and was once ranked nationally in the 99th percentile for microcystins by the NLA (U.S. EPA, 2009). Grand Lake was also ranked first in the state of Ohio for cyanobacterial bloom magnitude from 2003 through 2011 (Mishra et al., 2019). At Lake Jesup, where blooms have been noted throughout the year via satellite imagery (Coffer et al., 2020), average annual lake-scale bloom frequency reached 98% for 2019. Lake Elsinore, which reached an average annual lake-scale bloom frequency of 100%, had several documented lake-wide bloom events in 2019 with high levels of microcystins detected, exceeding 20 μg/L (California State Water Resources Control Board, 2020).

Previous studies or reports also support cyanobacterial blooms at Odell Lake (Oregon Health Authority, 2019a) and Upper Klamath Lake (Oregon Health Authority, 2019b) in Oregon, Lake Carmi in Vermont (Vermont, 2020), Island Lake in Nebraska (Efting et al., 2011), Utah Lake in Utah (Utah Department of Environmental Quality, 2019), Silver Lake in Iowa (Iqbal et al., 2006), Lake O' the Pines and Lake Palestine in Texas (Trevino and Petersen, 2020), Lakes Okeechobee (Rosen et al., 2017) and George (Srifa et al., 2016) in Florida, and Clear Lake in California (California State Water Resources Control Board, 2020; Kurobe et al., 2013).

3.3. Average annual state-scale cyanobacterial frequency

Average annual state-scale bloom frequencies for the year 2019 ranged from just over 1% in Georgia to nearly 65% in Oregon (Fig. 6).

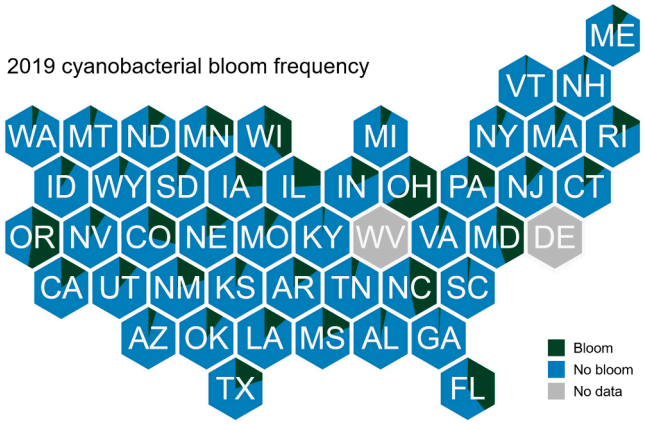


Fig. 6. Average annual state-scale cyanobacterial bloom frequencies for the year 2019. Each state is represented as a hexagon labeled with each state's two-letter state abbreviation. Each hexagon acts as a pie chart representing the proportion of pixels in the state whose annual frequency indicates a bloom is present above the detection limit of the sensor. West Virginia (WV) and Delaware (DE) do not have any lakes that are of sufficient size and shape to be resolvable at 300-m pixel resolution.

Just as pixel- and lake-scale bloom frequencies were weighted toward lower values, average state-scale bloom frequencies tended to be as well, with several states averaging relatively low frequency values. Of the 46 states with at least one resolvable lake, 21 states had an average annual bloom frequency below 10% for the year 2019, while Ohio and Oregon had average annual bloom frequencies above 50%.

Similar to average annual lake-scale frequencies, average annual state-scale frequencies suggested stability from 2017 through 2019 (Figures S7-S8). For each year, Oregon and Ohio consistently had the highest average annual bloom frequencies, while Wisconsin, North Carolina, Maryland, and Florida also exhibited a high average value relative to other states. Several states had low average annual bloom frequencies across years, and most resolvable pixels in the states of Vermont and Georgia even experienced close to no detectable blooms for the year 2018, with average annual state-scale bloom frequencies of 2% and 1%, respectively. As a note, Lake Champlain, which is monitored by the Vermont Department of Environmental Conservation and experiences cyanobacterial blooms (Shambaugh et al., 2019), is considered in state-scale summaries for New York rather than Vermont because of where its lake centroid falls. Results are also presented for the WHO risk levels (Figures S9-S11). Across nearly all states and years, the majority of detected blooms were above the WHO high-risk threshold (above 100,000 cells/mL), with lower proportions falling in the WHO moderate-risk level (between 20,000 and 10,000 cells/mL) and the WHO low-risk level (below 20,000 cells/mL). WHO risk level results are consistent with those presented in Clark et al. (2017).

3.4. Effect of time period used to compute cyanobacterial frequency

For the years 2017 through 2019, the beginning of each snow-free period representing no more than 10% snow and ice cover across CONUS ranged from 16 April to 21 April, and the end of each snow-free period ranged from 13 October to 26 November (Table S1). The longest snow-free period occurred in 2017 when 32 weekly composites were included, while the snow-free periods for 2018 and 2019 included the fewest at 26 and 25 weekly composites, respectively. The long snow-free period in 2017 is consistent with climatology, as 2017 was the third warmest year on record for the United States behind only 2012 and 2016, averaging 12.53 °C (NOAA, 2020). This study only used results for 2019 to compare annual frequencies to snow-free frequencies, but data are available from additional years to explore the stability of temporal biases presented here.

Average lake-scale bloom frequency was affected by the time period of choice (i.e., full-year versus a snow-free period) at three climate regions and not affected by the time period of choice at the remaining six climate regions (Table 1). In the Northwest Rockies and Plains, the difference between annual results and those computed for just the snow-free period was large (Wilcoxon signed-rank test effect size of 0.53). For this climate region, annual frequencies were lower than snow-free frequencies; however, this effect size is close to the recommended cutoff between a moderate effect and a large effect, which is 0.5. In the Ohio Valley and the Southeast, the difference between annual results and those computed for just the snow-free period was moderate (Wilcoxon signed-rank test effect size of 0.39 and 0.42, respectively). In the Ohio Valley, annual frequencies were slightly lower than snow-free frequencies. In the Southeast, however, annual frequencies were higher than snow-free frequencies. The remaining six climate regions indicated only a small difference in annual and snow-free frequencies (Wilcoxon signed-rank test effect sizes below 0.13), indicating the two sets of average lake-scale bloom frequencies had similar distributions.

Table 1
Results of the Wilcoxon signed-rank test comparing average annual lake-scale frequencies to average lake-scale frequencies computed using a subset of weekly composite to represent the snow-free period (defined as no more than 10% of CONUS surface area covered in snow and ice). Climate regions were defined according to Karl and Koss (1984) and n represents sample size.

Climate region	n	Annual median	Snow-free median	Effect size	Magnitude
Northeast	331	7.95%	7.87%	0.04	Small
Northwest	134	2.45%	1.43%	0.00	Small
Northwest Rockies and Plains	245	32.63%	36.01%	0.53	Large
Ohio Valley	91	11.62%	11.71%	0.39	Moderate
South	345	17.75%	16.11%	0.10	Small
Southeast	207	13.11%	3.98%	0.42	Moderate
Southwest	92	7.51%	5.32%	0.06	Small
Upper Midwest	655	11.35%	8.84%	0.08	Small
West	97	1.75%	0.83%	0.12	Small

Overall, these results indicate complex effects of removing cold-season data from calculations of annual bloom frequency, and do not unequivocally support the hypothesis of higher frequencies in summer months, as is commonly assumed. In fact, the difference between medians of the two average lake-scale frequencies was the largest for the Southeast; the median of annual frequencies was over 13% but dropped to less than 4% when considering the subset of snow-free weekly composites, indicative of cyanobacterial blooms throughout the winter months, a finding supported by Coffer et al. (2020). Southern states are also characterized by milder conditions than northern states throughout the year, and similar air temperatures are typically observed during southern latitude winters and northern latitude summers.

Caution should be taken in interpreting average annual cyanobacterial blooms frequencies between regions due to differing effects of snow and ice on the number of available observations, in addition to normal caveats such as Sun glint and cloud cover in these regions. Comparing average annual bloom frequencies to those derived using a snow-free period can help stakeholders understand potential biases introduced by the inclusion of cold-season data. Cold-season data are particularly affected by missing observations due to snow and ice cover, and blooms tend to be less prevalent in the cold-season across the United States, with blooms peaking in late summer and early fall across most of CONUS (Coffer et al., 2020). These findings indicate that, at a broad scale, bloom frequency results are only mildly affected by the choice of either full-year observations or a subset of data to exclude cold-season observations. Exceptions include the Northwest Rockies and Plains and the Ohio Valley, where annual observations tended toward lower values than snow-free observations, and the Southeast, where annual observations tended toward higher values than snow-free observations, indicative of cyanobacterial blooms throughout the winter months.

3.5. Spatial patterns in cyanobacterial frequency

In the United States, several lakes have been known to have persistent cyanobacterial blooms, including Grand Lake in Ohio (Jacquemin et al., 2018; Mishra et al., 2019), Lake Apopka in Florida (Clark et al., 2017; Mishra et al., 2019), and Lake Champlain along the border of Vermont and New York (Isles et al., 2015); however, no national scale analyses have been conducted to identify clusters of lakes with high or low frequencies of cyanobacterial blooms. The results of this cluster analysis provide quantitative evidence for spatial patterns that may be visually discerned based on the average annual lake-scale cyanobacterial frequencies presented in Fig. 4, and act as the first national analysis of spatial clustering for cyanobacterial frequency in large lakes. The Getis-Ord Gi* statistic does not quantify the number of clusters present or specify which cluster a given lake falls within. Nevertheless, based on average annual lake-scale cyanobacterial bloom frequencies for the year 2019, results indicate several spatial clusters of both high and low cyanobacterial bloom frequency across CONUS (Fig. 7). Perhaps more importantly, spatial patterns presented here clearly indicate that clusters transcend state, watershed, and eco-regional boundaries.

A lake was characterized as part of a cluster of high cyanobacterial bloom frequencies if it had a large, positive z-score and a small p-value; 568 lakes met these criteria with 99% confidence and an additional 193

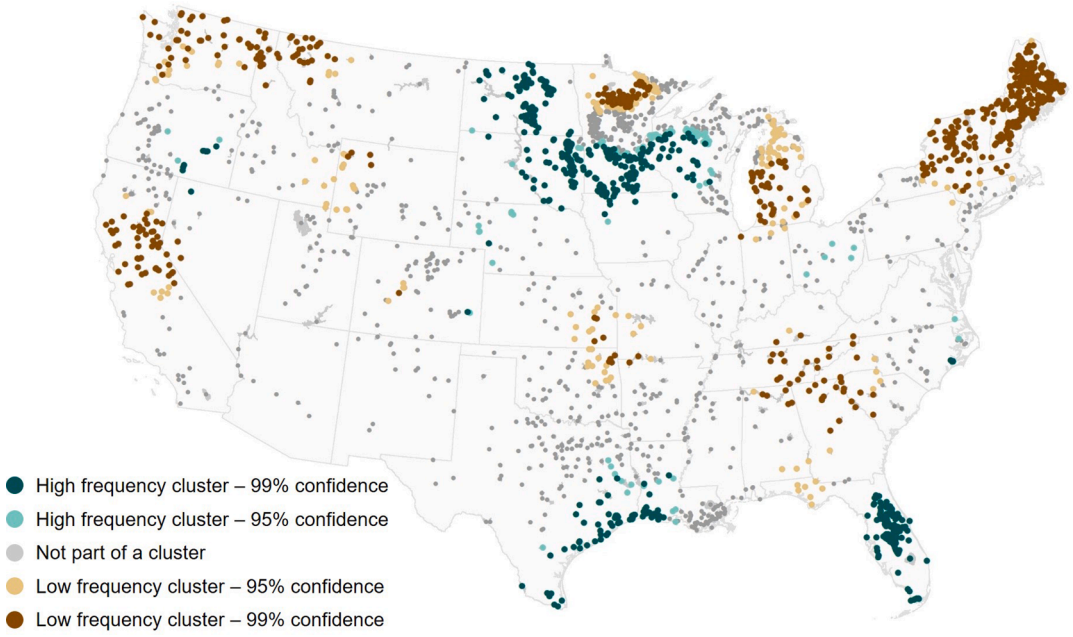


Fig. 7. Results of the Getis-Ord Gi* statistic applied to average annual lake-scale cyanobacterial bloom frequencies for the year 2019. This spatial clustering algorithm considers both spatial location and the corresponding data values.

lakes met these criteria with 95% confidence. A lake was characterized as part of a cluster of low cyanobacterial bloom frequencies if it had a large, negative z-score and a small p-value; 491 lakes met these criteria with 99% confidence and an additional 97 lakes met these criteria with 95% confidence. The 847 remaining lakes did not correspond to a spatial cluster of cyanobacterial frequency with a confidence of at least 95%. As a reminder, 300-m satellite imagery can only resolve relatively large lakes across CONUS, so spatial clusters may exist among lakes not observable with OLCI imagery due to pixel size limitations. Additionally, limitations of the bloom frequency metric, including missing data due to Sun glint, cloud cover, and snow and ice, can affect the average annual lake-scale summaries used in this analysis which can thus affect the resulting cluster analysis.

A high frequency cluster can be found spanning North Dakota, South Dakota, southern Minnesota, and Wisconsin. The Gulf Coast in Texas and Louisiana contained a high frequency cluster that covers a smaller spatial extent. Nearly all lakes in Florida were included in a high frequency cluster with 99% confidence. Lakes throughout much of New England including Maine, Vermont, New Hampshire, Massachusetts, Rhode Island, Connecticut, and New York constituted a low frequency cluster. Low frequency clusters also existed in the following areas: the Pacific Northwest including Washington, northern Idaho, and western Montana; central California; northern Minnesota; Michigan; the Central Plains; and parts of the mid-Atlantic including Tennessee, western North Carolina, South Carolina, northern Georgia, and northern Alabama.

Spatial clustering algorithms have been used in previous studies to indicate areas of increased cyanobacterial presence, but on a much smaller spatial scale within specific waterbodies. For example, Yunus et al. (2015) identified groups of pixels with high chlorophyll-a occurrence using data collected from Landsat-8 in Tokyo Bay in Japan. Gaskill and Woller-Skar (2018) used field measurements at Little Traverse Lake in Michigan to identify spatial clusters of microcystis blooms. In Chile, spatial statistics have been used to identify spatial clustering of cyanobacterial communities (Warren-Rhodes et al., 2007). The results of the clustering algorithm presented here are intended for a broader application with the potential to identify state- or regional-scale areas in which cyanobacterial blooms persist. Additionally, while out of the scope of this study, results from this cluster analysis can be used in combination with additional data products to understand large-scale drivers of cyanobacterial blooms that have been previously reported at the regional level (e.g., Myer et al., 2020).

3.6. Limitations

Several limitations exist in the quantification of pixel-scale bloom frequencies and both lake- and state-scale summaries. Retrievals from a passive satellite sensor can only consider contributions from aquatic constituents present to a depth of about 2 m in clear waters in the red region of the electromagnetic spectrum used for characterizing cyanobacteria (Mishra et al., 2005) and much less than 2 m in more turbid waters (Wynne et al., 2010). Cyanobacterial blooms, however, can exist at many depths throughout the water column and have the ability to regulate their buoyancy in an effort to find more ideal environmental conditions (Bartram and Chorus, 1999); thus, *in situ* observation networks at depth remain essential. Graham et al. (2008) described six potential water column distributions of cyanobacteria: (1) shoreline, near-shore, and open water accumulations and scums, (2) even distribution throughout the photic zone or epilimnion, (3) specific depth in the photic zone, (4) metalimnetic bloom, (5) even distribution throughout the entire water column, and (6) under ice bloom. Satellite remote sensing has the potential to characterize cyanobacterial blooms in four of these scenarios; ocean color satellites are unable to retrieve information about the water column in the presence of ice cover, and a metalimnetic bloom could be at a depth the satellite signal does not reach, but this scenario was noted as more of a concern for drinking water supplies. While cyanobacterial blooms can exist at many depths within the water column, recreational exposure, at least, is primarily limited to the surface. This study did not consider satellite-retrieved water transparency, which could be combined with satellite-retrieved cyanobacterial abundance to estimate the maximum depth of the satellite signal. Information quantifying the depth of the satellite signal would provide insight regarding the portion of the water column being characterized by the cyanobacterial frequency metric.

Results presented here are based on 300-m satellite imagery, which has limitations at the lake- and state-scales as both areas within each resolvable lake are excluded from analysis, including narrow reaches and pixels along the land/water interface, and lakes within each state are excluded from analysis if they cannot accommodate at least one,

300-m water-only pixel after applying the exclusion criteria described in Section 2.1. Schaeffer and Myer (2020) assessed the coverage of estuaries and sub-estuaries across the United States with various spatial resolutions. Satellite platforms with higher spatial resolution could improve both lake- and state-scale characterizations. One such satellite sensor is ESA's Sentinel-2 satellite platform. However, while chlorophyll-specific algorithms have been developed for Sentinel-2 imagery (Gilerson et al., 2010; Pahlevan et al., 2020), cyanobacteria-specific algorithms have not been demonstrated for large-scale assessments because of limitations in the spectral capabilities of the satellite instrumentation. We did not investigate if results presented here can be interpolated to smaller lakes, although satellite imagery offering a finer spatial resolution could help address this limitation. And while higher spatial resolution imagery can improve data coverage, additional limitations can emerge. For example, for large scale data processing, automated pixel flagging for non-water objects such as boats, bridges, marinas, and even migratory birds would be required as the average reflectance within a finer spatial resolution pixel can be dominated by such artifacts, a limitation less problematic in coarser, 300-m satellite imagery.

Potential biases resulting from missing data due to snow and ice extent in the winter months were explored, but other limitations in satellite data collection exist that could result in problematic comparisons when considering lakes or states in different regions. Cloud cover, for example, reduces temporal coverage, and comparisons across states with differing patterns in cloud cover could be biased. An analysis such as that presented here for snow and ice could be applied to cloud cover, but the assumptions are less straightforward because the spatial and temporal distribution of cloud cover can be more varied. An additional limitation of satellite image acquisition whose potential biases on annual cyanobacterial frequency metrics was not explored is Sun glint, which hinders data collection, particularly in the summer months and at lower latitudes due to changes in solar elevation.

This study focuses on the temporal frequency of cyanobacterial blooms and does not characterize cyanobacterial spatial extent (Urquhart et al., 2017), occurrence (Coffer et al., 2020), nor magnitude (Mishra et al., 2019), which have been explored in previous studies. Data are quality flagged and discarded in the presence of Sun glint, cloud cover, or snow and ice. The snow and ice mask presented in Urquhart and Schaeffer (2020) is likely more conservative than is needed but currently offers the most comprehensive and consistent product available. This study demonstrated that eliminating pixels during the cold season because of snow and ice cover can create bias in certain climate regions. Additionally, cyanobacterial blooms have been noted under ice, including some blooms containing cyanotoxins (Bertilsson et al., 2013; Hampton et al., 2017; Üveges et al., 2012; Wejnerowski et al., 2018).

4. Applications

This study provides the first national scale application of a cyanobacterial frequency metric introduced at a regional scale in Clark et al. (2017). Source code, documentation, and results accompanying this manuscript can be found at doi:10.23719/1520731. The methods presented here can be adjusted to categorize different cyanobacterial concentrations (e.g., WHO risk thresholds), depending on stakeholder needs and requirements. Moreover, while annual frequencies were presented here, the same methods can be altered to compute seasonal or monthly frequencies as well. Lake- and state-scale frequency results may assist in the prioritization of sampling resources and mitigation efforts. Seasonal comparisons may provide guidance on the interpretation of average annual bloom frequencies, indicating regions in which annual results may be slightly biased by the inclusion of cold-season data.

Pixel-scale frequencies can assist in identifying segments of a lake that are more prone to cyanobacterial blooms, which can be of particular interest if the lake is used for recreational activities or if it contains a drinking water intake. A successful use case of cyanobacterial bloom frequency results presented here occurred in Salem, Oregon. In 2018, vulnerable residents in the city of Salem were under a "Do Not Drink" advisory for several weeks after elevated levels of cyanotoxins were detected in finished drinking water (The Novak Consulting Group, 2018). Preliminary bloom frequency results were presented to the city of Salem for Detroit Lake, upstream from the intake, to identify areas of high cyanobacterial frequency and prioritize sampling.

Moreover, many studies have claimed an increase in the frequency of cyanobacterial blooms (Anderson et al., 2002), but few studies have actually quantified such changes across the United States. While this study does not offer a trend analysis, it does present a consistent metric that can be used to assess changes in cyanobacterial frequency as a longer time series becomes available with ongoing satellite data collection within the limitations specified earlier. Assuming availability of the full constellation of Sentinel-3 satellites, at least 25 years of imagery are expected to be available since the launch of Sentinel-3A in 2016 and can be used to improve our ability to quantify trends in cyanobacterial frequency.

CRediT authorship contribution statement

Megan M. Coffer: Conceptualization, Methodology, Software, Formal analysis, Validation, Visualization, Writing - original draft. **Blake A. Schaeffer:** Conceptualization, Methodology, Supervision, Resources, Project administration, Funding acquisition, Writing - review & editing. **Wilson B. Salls:** Conceptualization, Writing - review & editing. **Erin Urquhart:** Conceptualization, Methodology, Writing - review & editing. **Keith A. Loftin:** Funding acquisition, Conceptualization, Writing - review & editing. **Richard P. Stumpf:** Funding acquisition, Conceptualization, Writing - review & editing. **P. Jeremy Werdell:** Funding acquisition, Conceptualization, Writing - review & editing. **John A. Darling:** Conceptualization, Methodology, Supervision, Project administration, Writing - review & editing.

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.

Acknowledgements

The authors thank C. Crawford, J. Carter, M. Fry, and H. Golden for their contributions in manuscript review and an anonymous reviewer for their assistance in improving the communication of the study. This work was supported by the NASA Ocean Biology and Biogeochemistry Program/Applied Sciences Program (proposals 14-SMDUNSOL14-0001 and SMDSS20-000) and by U.S. EPA, NOAA, U.S. Geological Survey Toxic Substances Hydrology Program, and Oak Ridge Institute for Science and Technology (ORISE). This article has been reviewed by the CEMM and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2021.107822.

References

Anderson-Abbs, B., Howard, M., Taberski, K., Worcester, K., 2016. California freshwater harmful algal blooms assessment and support strategy. California State Water Resources Control Board. URL: https://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/925_CaliforniaFreshwaterHABAssessment.pdf.

Anderson, D.M., Glibert, P.M., Burkholder, J.M., 2002. Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuaries 25, 704–726. https://doi.org/10.1007/BF02804901.
Bartram, J., Chorus, I., 1999. Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring, and Management, 1st ed. CRC Press, London. https://doi.org/10.1201/9781482295061.
Berger, P., Brooks, J., Evens, T., Gobler, C., Graham, J., Berger, P., Brooks, J., Evens, T., Graham, J., Hyde, J., Karner, D., O'Shea, D.K., Paul, V., Paerl, H., Piehler, M., Rosen, B., Santelmann, M., Tester, P., Westrick, J., 2008. Cyanobacterial Harmful Algal Blooms. In: Chapter, 9. Causes, Prevention, and Mitigation Workgroup Report.
Bertilsson, S., Burgin, A., Carey, C.C., Fey, S.B., Grossart, H.-P., Grubisic, L.M., Jones, I. D., Kirillin, G., Lennon, J.T., Shade, A., Smyth, R.L., 2013. The under-ice microbiome of seasonally frozen lakes. Limnol. Oceanogr. 58, 1998–2012. https://doi.org/10.4319/lo.2013.58.6.1998.
Beutel, M.W., Horne, A.J., 2018. Nutrient fluxes from profundal sediment of ultra-oligotrophic Lake Tahoe, California/Nevada: implications for water quality and management in a changing climate. Water Resour. Res. 54, 1549–1559. https://doi.org/10.1002/2017WR020907.
California State Water Resources Control Board, 2020. Surface Water - Freshwater Harmful Algal Blooms.
Clark, J.M., Schaeffer, B.A., Darling, J.A., Urquhart, E.A., Johnston, J.M., Ignatius, A.R., Myer, M.H., Loftin, K.A., Werdell, P.J., Stumpf, R.P., 2017. Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. Ecol. Indic. 80, 84–95. https://doi.org/10.1016/j.ecolind.2017.04.046.
Coats, R., Lewis, J., Alvarez, N., Arneson, P., 2016. Temporal and spatial trends in nutrient and sediment loading to lake tahoe, California-Nevada, USA. JAWRA J. Am. Water Resour. Assoc. 52, 1347–1365. https://doi.org/10.1111/1752-1688.12461.
Coffer, M.M., Schaeffer, B.A., Darling, J.A., Urquhart, E.A., Salls, W.B., 2020. Quantifying national and regional cyanobacterial occurrence in US lakes using satellite remote sensing. Ecol. Indic. 111, 105976 https://doi.org/10.1016/j.ecolind.2019.105976.
Cohen, J., 1992. A power primer. Psychol. Bull. 112, 155. https://doi.org/10.1037//0033-2909.112.1.155.
Dubrovsky, N.M., Graham, J.L., Eberts, S.M., 2016. Cyanobacterial Harmful Algal Blooms and U.S. Geological Survey Science Capabilities View of a cyanobacterial bloom in Milford Lake. https://doi.org/10.3133/ofr20161174.
Efting, A.A., Snow, D.D., Fritz, S.C., 2011. Cyanobacteria and microcystin in the Nebraska (USA) Sand Hills Lakes before and after modern agriculture. J. Paleolimnol. 46, 17–27. https://doi.org/10.1007/s10933-011-9511-3.
ESRI, 2019. ArcGIS Desktop: Release 10.7.1.
Gaskill, J.A., Woller-Skar, M.M., 2018. Do invasive dreissenid mussels influence spatial and temporal patterns of toxic *Microcystis aeruginosa* in a low-nutrient Michigan lake? Lake Reserv. Manag. 34, 244–257. https://doi.org/10.1080/10402381.2018.1423587.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24, 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.
Gilerson, A.A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Opt. Express 18, 24109–24125. https://doi.org/10.1364/OE.18.024109.
GLCR, 2013. Microcystis Outbreak Closes Lake Erie Water Treatment Plant. http://greatlakesresilience.org/climate-environment/microcystis-outbreak-closes-lake-erie-water-treatment-plant-0.
Gons, H.J., Hakvoort, H., Peters, S.W.M., Simis, S.G.H., 2005. Optical Detection of Cyanobacterial Blooms. In: Huisman, J., Matthijs, H.C.P., Visser, P.M. (Eds.), Harmful Cyanobacteria. Springer, Netherlands, Dordrecht, pp. 177–199. https://doi.org/10.1007/1-4020-3022-3_8.
Graham, J.L., Loftin, K.A., Ziegler, A.C., Meyer, M.T., 2008. Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs. U.S. Geological Survey. URL: https://pubs.usgs.gov/sir/2008/5038/pdf/SIR2008-5038.pdf.
Hampton, S.E., Galloway, A.W.E., Powers, S.M., Ozersky, T., Woo, K.H., Batt, R.D., Labou, S.G., O'Reilly, C.M., Sharma, S., Lottig, N.R., Stanley, E.H., North, R.L., Stockwell, J.D., Adrian, R., Weyhenmeyer, G.A., Arvola, L., Baulch, H.M., Bertani, I., Bowman, L.L., Carey, C.C., Catalan, J., Colom-Montero, W., Domine, L.M., Felip, M., Granados, I., Gries, C., Grossart, H.-P., Haberman, J., Haldna, M., Hayden, B., Higgins, S.N., Jolley, J.C., Kahilainen, K.K., Kaup, E., Kehoe, M.J., MacIntyre, S., Mackay, A.W., Mariash, H.L., McKay, R.M., Nixdorf, B., Nõges, P., Nõges, T., Palmer, M., Pierson, D.C., Post, D.M., Pruett, M.J., Rautio, M., Read, J.S., Roberts, S. L., Rücker, J., Sadro, S., Silow, E.A., Smith, D.E., Sterner, R.W., Swann, G.E.A., Timofeyev, M.A., Toro, M., Twiss, M.R., Vogt, R.J., Watson, S.B., Whiteford, E.J., Xenopoulos, M.A., 2017. Ecology under lake ice. Ecol. Lett. 20, 98–111. https://doi.org/10.1111/ele.12699.
Hu, C., Lee, Z., Ma, R., Yu, K., Li, D., Shang, S., 2010. Moderate Resolution Imaging Spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake. China. J. Geophys. Res. Ocean. 115 https://doi.org/10.1029/2009JC005511.
Iqbal, M.Z., Brown, E.J., Clayton, M.E., 2006. Distribution of phosphorus in a biologically restricted lake in Iowa. USA. J. Hydrol. 326, 349–366. https://doi.org/10.1016/j.jhydrol.2005.11.006.
Isles, P.D.F., Giles, C.D., Gearhart, T.A., Xu, Y., Druschel, G.K., Schroth, A.W., 2015. Dynamic internal drivers of a historically severe cyanobacteria bloom in Lake Champlain revealed through comprehensive monitoring. J. Great Lakes Res. 41, 818–829. https://doi.org/10.1016/j.jglr.2015.06.006.
Jacquemin, S.J., Johnson, L.T., Dirksen, T.A., McGlinch, G., 2018. Changes in water quality of Grand Lake St. Marys watershed following implementation of a distressed watershed rules package. J. Environ. Qual. 47, 113–120. https://doi.org/10.2134/jeq2017.08.0338.
Kahru, M., Savchuk, O.P., Elmgren, R., 2007. Satellite measurements of cyanobacterial bloom frequency in the Baltic Sea: interannual and spatial variability. Mar. Ecol. Prog. Ser. 343, 15–23. https://doi.org/10.3354/meps06943.
Karl, T., Koss, W.J., 1984. Regional and National Monthly, Seasonal, and Annual Temperature Weighted by Area, 1895–1983, Historical Climatology Series 4–3. Asheville, NC.
Kassambara, A., 2020. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. URL: https://cran.r-project.org/package=rstatix.
Kurobe, T., Baxa, D. V, Mioni, C.E., Kudela, R.M., Smythe, T.R., Waller, S., Chapman, A. D., Teh, S.J., 2013. Identification of harmful cyanobacteria in the Sacramento-San Joaquin Delta and Clear Lake, California by DNA barcoding. Springerplus 2, 491. https://doi.org/10.1186/2193-1801-2-491.
Kutser, T., 2009. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. Int. J. Remote Sens. 30, 4401–4425. https://doi.org/10.1080/01431160802562305.
Lunetta, R.S., Schaeffer, B.A., Stumpf, R.P., Keith, D., Jacobs, S.A., Murphy, M.S., 2015. Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. Remote Sens. Environ. 157, 24–34. https://doi.org/10.1016/j.rse.2014.06.008.
Mishra, D.R., Narumalani, S., Rundquist, D., Lawson, M., 2005. Characterizing the vertical diffuse attenuation coefficient for downwelling irradiance in coastal waters: Implications for water penetration by high resolution satellite data. ISPRS J. Photogramm. Remote Sens. 60, 48–64. https://doi.org/10.1016/j.isprsjprs.2005.09.003.
Mishra, S., Stumpf, R.P., Schaeffer, B.A., Werdell, P.J., Loftin, K.A., Meredith, A., 2021. Evaluation of a satellite-based cyanobacteria bloom detection algorithm using field-measured Microcystin data. Sci. Total Environ. 774. https://doi.org/10.1016/j.scitotenv.2021.145462.
Mishra, S., Stumpf, R.P., Schaeffer, B.A., Werdell, P.J., Loftin, K.A., Meredith, A., 2019. Measurement of cyanobacterial bloom magnitude using satellite remote sensing. Sci. Rep. 9, 18310. https://doi.org/10.1038/s41598-019-54453-y.
Myer, M.H., Urquhart, E., Schaeffer, B.A., Johnston, J.M., 2020. Spatio-temporal modeling for forecasting high-risk freshwater cyanobacterial harmful algal blooms in Florida. Front. Environ. Sci. 8 https://doi.org/10.3389/fenvs.2020.581091.
NASA JPL, 2013. NASA Shuttle Radar Topography Mission Water Body Data Shapefiles & Raster Files. NASA EOSDIS Land Processes, DAAC Sioux Falls, SD, USA.
National Ice Center, 2008. IMS Daily Northern Hemisphere Snow and Ice Analysis at 1 km, 4 km, and 24 km Resolutions, Version 1. https://doi.org/10.7265/N52R3PMC.
NCDC, 2019. State of the Climate: National Climate Report for Annual 2019. NOAA National Center for Environmental Information. URL: https://www.ncdc.noaa.gov/sotc/national/201913.
ND DEQ, 2019. Harmful Algal Bloom Water Advisories/Warnings. North Dakota Department of Environmental Quality. URL: https://deq-ndgov.maps.arcgis.com/apps/Shortlist/index.html?appid=9b28a6b198f24847be68742d3eb5b927.
NOAA, 2020. Climate at a Glance: National Time Series. NOAA National Center for Environmental Information. URL: https://www.ncdc.noaa.gov/cag/.
NOAA NCEI, 2021. U.S. Billion-Dollar Weather and Climate Disasters. https://doi.org/10.25921/stkw-7w73.
Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. 27, 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x.
Oregon Health Authority, 2019a. Odell Lake Recreational use health advisory lifted August 14 [WWW Document]. URL https://www.oregon.gov/oha/ERD/Pages/2019-August-Odell-Lake-Health-Advisory-Lift.aspx.
Oregon Health Authority, 2019b. Upper Klamath Lake recreational use health advisory lifted October 3 [WWW Document]. URL https://www.oregon.gov/oha/ERD/Pages/2019-October-Upper-Klamath-Lake-Recreational-Health-Advisory-Lifted.aspx.
Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. Remote Sens. Environ., 111604 https://doi.org/10.1016/j.rse.2019.111604.
Papenfus, M., Schaeffer, B., Pollard, A.I., Loftin, K., 2020. Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. Environ. Monit. Assess. 192, 808. https://doi.org/10.1007/s10661-020-08631-5.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. Url: http://www.r-project.org/.
Reynolds, C.S., 2006. Ecology of Phytoplankton, Ecology, Biodiversity, and Conservation. Cambridge University Press, Cambridge.
Rosen, B.H., Davis, T.W., Gobler, C.J., Kramer, B.J., Loftin, K.A., Survey, U.S.G., 2017. Cyanobacteria of the 2016 Lake Okeechobee and Okeechobee Waterway harmful algal bloom, U.S. Geological Survey Open-File Report 2017-1054. Reston, VA. https://doi.org/10.3133/ofr20171054.
Schaeffer, B., Myer, 2020. Resolvable estuaries for satellite derived water quality within the continental United States. Remote Sens. Lett. 11 (6), 535–544. https://doi.org/10.1080/2150704X.2020.1717013.
Schaeffer, B.A., Bailey, S.W., Conmy, R.N., Galvin, M., Ignatius, A.R., Johnston, J.M., Keith, D.J., Lunetta, R.S., Parmar, R., Stumpf, R.P., Urquhart, E.A., Werdell, P.J., Wolfe, K., 2018. Mobile device application for monitoring cyanobacteria harmful algal blooms using Sentinel-3 satellite Ocean and Land Colour Instruments. Environ. Model. Softw. 109, 93–103. https://doi.org/10.1016/j.envsoft.2018.08.015.
Shambaugh, A., O'Brien, B., Fisher, L., Campbell, H., 2019. Cyanobacteria Monitoring on Lake Champlain Summer 2018. URL: https://link-springer-com.prox.lib.ncsu.edu/content/pdf/10.1007%2F978-0-387-75865-7.pdf.

Śliwińska-Wilczewska, S., Maculewicz, J., Barreiro Felpeto, A., Latała, A., 2018. Allelopathic and bloom-forming picocyanobacteria in a changing world. Toxins (Basel). 10, 48. https://doi.org/10.3390/toxins10010048.

Srifa, A., Phlips, E.J., Cichra, M.F., Hendrickson, J.C., 2016. Phytoplankton dynamics in a subtropical lake dominated by cyanobacteria: cyanobacteria 'Like it Hot' and sometimes dry. Aquat. Ecol. 50, 163–174. https://doi.org/10.1007/s10452-016-9565-4.

Stroming, S., Robertson, M., Mabee, B., Kuwayama, Y., Schaeffer, B., 2020. Quantifying the Human Health Benefits of Using Satellite Information to Detect Cyanobacterial Harmful Algal Blooms and Manage Recreational Advisories in U.S. Lakes. GeoHealth 4, e2020GH000254. https://doi.org/10.1029/2020GH000254.

Stumpf, R.P., Davis, T.W., Wynne, T.T., Graham, J.L., Loftin, K.A., Johengen, T.H., Gossiaux, D., Palladino, D., Burtner, A., 2016. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. Harmful Algae 54, 160–173. https://doi.org/10.1016/j.hal.2016.01.005.

The Novak Consulting Group, 2018. City of Salem Water Advisory After-Action Assessment. URL: https://www.cityofsalem.net/CityDocuments/water-advisory-after-action-report-2018.pdf.

Trevino, J., Petersen, C.B., 2020. Assessment of Field and Laboratory Methods for the Detection and Analyses of Cyanobacteria and Cyanotoxins in Texas Reservoirs. https://doi.org/10.5066/P9ASYKM7.

U.S. EPA, 2016. National Lakes Assessment 2012: A Collaborative Survey of Lakes in the United States. EPA 841-R-16-113. Washington, DC. Url: https://nationallakesassessment.epa.gov/.

U.S. EPA, 2009. National lakes assessment: a collaborative survey of the nation's lakes.

Urquhart, E.A., Schaeffer, B.A., 2020. Envisat MERIS and Sentinel-3 OLCI satellite lake biophysical water quality flag dataset for the contiguous United States. Data Br. 28, 104826. https://doi.org/10.1016/j.dib.2019.104826.

Urquhart, E.A., Schaeffer, B.A., Stumpf, R.P., Loftin, K.A., Werdell, P.J., 2017. A method for examining temporal changes in cyanobacterial harmful algal bloom spatial extent using satellite remote sensing. Harmful Algae 67, 144–152. https://doi.org/10.1016/J.HAL.2017.06.001.

Utah Department of Environmental Quality, 2019. Utah Lake Algal Bloom Monitoring 2019 [WWW Document]. URL https://deq.utah.gov/water-quality/utah-lake-algal-bloom-monitoring-2019.

Üveges, V., Tapolczai, K., Krienitz, L., Padisák, J., 2012. Photosynthetic characteristics and physiological plasticity of an Aphanizomenon flos-aquae (Cyanobacteria, Nostocaceae) winter bloom in a deep oligo-mesotrophic lake (Lake Stechlin, Germany). In: Phytoplankton Responses to Human Impacts at Different Scales. Springer, Netherlands, pp. 263–272. https://doi.org/10.1007/978-94-007-5790-5_20.

Vermont DEC, 2020. 2019 Lake Carmi Clean Water Progress Report. Vermont Agency of Natural Resources' Department of Environmental Conservation (DEC). URL: https://dec.vermont.gov/sites/dec/files/wsm/lakes/docs/2019 Lake Carmi Clean Water Progress Report FINAL.pdf.

Warren-Rhodes, K.A., Dungan, J.L., Piatek, J., Stubbs, K., Gómez-Silva, B., Chen, Y., McKay, C.P., 2007. Ecology and spatial pattern of cyanobacterial community island patches in the Atacama Desert. Chile. J. Geophys. Res. Biogeosciences 112. https://doi.org/10.1029/2006JG000305.

Wejnerowski, Ł., Rzymski, P., Kokociński, M., Meriluoto, J., 2018. The structure and toxicity of winter cyanobacterial bloom in a eutrophic lake of the temperate zone. Ecotoxicology 27, 752–760. https://doi.org/10.1007/s10646-018-1957-x.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin. https://doi.org/10.2307/3001968.

World Health Organization, 2003. Algae and Cyanobacteria in Fresh Water, in: Guidelines for Safe Recreational Water Environments. pp. 136–158.

Wynne, T.T., Meredith, A., Briggs, T., Litaker, W., Stumpf, R.P., 2018. Harmful algal bloom forecasting branch ocean color satellite imagery processing guidelines. NOAA National Centers for Coastal Ocean Science. URL: https://repository.library.noaa.gov/view/noaa/20270.

Wynne, T.T., Stumpf, R., Tomlinson, M., Dyble, J., 2010. Characterizing a cyanobacterial bloom in Western Lake Erie using satellite imagery and meteorological data. Limnol. Oceanogr. 55, 2025–2036. https://doi.org/10.4319/lo.2010.55.5.2025.

Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., Warner, R.A., Tester, P.A., Dyble, J., Fahnenstiel, G.L., 2008. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. Int. J. Remote Sens. 29, 3665–3672. https://doi.org/10.1080/01431160802007640.

Yang, Z., Zhang, M., Shi, X., Kong, F., Ma, R., Yu, Y., 2016. Nutrient reduction magnifies the impact of extreme weather on cyanobacterial bloom formation in large shallow Lake Taihu (China). Water Res. 103, 302–310. https://doi.org/10.1016/j.watres.2016.07.047.

Yunus, A.P., Dou, J., Sravanthi, N., 2015. Remote sensing of chlorophyll-a as a measure of red tide in Tokyo Bay using hotspot analysis. Remote Sens. Appl. Soc. Environ. 2, 11–25. https://doi.org/10.1016/j.rsase.2015.09.0.

Zanchett, G., Oliveira-Filho, E.C., 2013. Cyanobacteria and cyanotoxins: from impacts on aquatic ecosystems and human health to anticarcinogenic effects. Toxins (Basel) 5, 1896–1917. https://doi.org/10.3390/toxins5101896.