Published in final edited form as:
Sci Total Environ. 2023 April 15; 869: 161784. doi:10.1016/j.scitotenv.2023.161784.

Identifying lakes at risk of toxic cyanobacterial blooms using satellite imagery and field surveys across the United States

Amalia M. Handler[1,*], **Jana E. Compton**[1], **Ryan A. Hill**[1], **Scott G. Leibowitz**[1], **Blake A. Schaeffer**[2]

[1]Center for Public Health and Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency; Corvallis, OR, 97333.

[2]Center for Environmental Measurement and Modeling, Office of Research and Development, U.S. Environmental Protection Agency; Durham, NC, 27711.

Abstract

Harmful algal blooms caused by cyanobacteria are a threat to global water resources and human health. Satellite remote sensing has vastly expanded spatial and temporal data on lake cyanobacteria, yet there is still acute need for tools that identify which waterbodies are at-risk for toxic cyanobacterial blooms. Algal toxins cannot be directly detected through imagery but monitoring toxins associated with cyanobacterial blooms is critical for assessing risk to the environment, animals, and people. The objective of this study is to address this need by developing an approach relating satellite imagery on cyanobacteria with field surveys to model the risk of toxic blooms among lakes. The Medium Resolution Imaging Spectrometer (MERIS) and United States (US) National Lakes Assessments are leveraged to model the probability among lakes of exceeding lower and higher demonstration thresholds for microcystin toxin, cyanobacteria, and chlorophyll *a*. By leveraging the large spatial variation among lakes using two national-scale data sources, rather than focusing on temporal variability, this approach avoids many of the previous challenges in relating satellite imagery to cyanotoxins. For every satellite-derived lake-level Cyanobacteria Index (CI_cyano) increase of 0.01 CI_cyano/km^2, the odds of exceeding six bloom thresholds increased by 23–54%. When the models were applied to the 2,192 satellite monitored lakes in the US, the number of lakes identified with ≥75% probability of exceeding the thresholds included as many as 335 lakes for the lower thresholds and 70 lakes for the higher thresholds, respectively. For microcystin, the models identified 162 and 70 lakes with ≥75% probability of exceeding the lower (0.2 μg/L) and higher (1.0 μg/L) thresholds, respectively. This approach represents a critical advancement in using satellite imagery and field data to identify lakes at risk

[*]Corresponding author. handler.amalia@epa.gov.

Author contributions: Each author's contribution(s) are as follows:
Conceptualization: AMH, JEC, SGL, RAH
Methodology: AMH, BAS, RAH
Investigation: AMH
Writing – original draft: AMH
Writing – review & editing: AMH, JEC, BAS, RAH, SGL

Competing interests: Authors declare that they have no competing interests.

Supplementary Materials
Text S1
Figs. S1 to S5
Tables S1 to S2

for developing toxic cyanobacteria blooms. Such models can help translate satellite data to aid water quality monitoring and management.

Keywords

Harmful algal blooms; cyanobacteria; microcystin; remote sensing; lakes

1 Introduction

Harmful algal blooms caused by cyanobacteria (cyanoHABs) are a pressing concern for waterbodies that provide drinking water, recreational opportunities, and wildlife habitat (Paerl et al., 2001). CyanoHABs are problematic because this type of algae can produce cyanotoxins that can harm people, pets, livestock, and wildlife through water contact or consumption (Paerl et al., 2001). The most commonly detected cyanotoxins are microcystins, a group of many toxic compounds that can cause skin irritation and serious illness or, in rare cases, death (Beaver et al., 2014; Paerl et al., 2001; World Health Organization, 2003). In 2012, the proportion of lakes in the United States (US) with detectable microcystin was 39% and the population of lakes with high levels of cyanobacteria grew by 8.3% from 2007 (US EPA, 2016). The factors driving cyanoHABs are varied and context-specific; however, there is growing consensus that climate change is expected to increase the incidence of cyanoHABs (Chapra et al., 2017; Elliott, 2012; Huo et al., 2019; O'Neil et al., 2012; Rigosi et al., 2015). Given the threat of toxin-producing cyanoHABs, water resource managers and communities reliant on these waterbodies need better tools for prediction and monitoring.

Monitoring cyanoHABs is critical for managing risk to public health; however, a major challenge is acquiring adequate data. Blooms can develop and change quickly over days or even hours. Monitoring cyanoHABs in the field is time- and resource-intensive. Field monitoring generally yields limited spatial and temporal information about blooms. Large amounts of data are required to understand dynamics and anticipate a cyanoHAB event within a given lake. Information about the drivers of cyanoHABs add to the already intensive data requirements. Many variables are thought to drive cyanoHABs including temperature (Chapra et al., 2017; Rigosi et al., 2015), excess nutrient availability (O'Neil et al., 2012), and hydrology with respect to the sources of water to a lake (Brookfield et al., 2021; Stumpf et al., 2016b), water residence time (Xin et al., 2020; Xu et al., 2021), and mixing behavior (Taranu et al., 2012). The high data demands of studying the dynamics of cyanoHABs make developing predictive models challenging. Indeed, a small number of intensively sampled lakes that have experienced drinking water disruptions from cyanoHABs account for many of the developed forecasting models (Rousso et al., 2020). However, predictive models developed for site-specific cyanoHABs can rarely be applied to other lakes (Rousso et al., 2020; Taranu et al., 2012). Given the massive number of waterbodies across the globe that are potentially at-risk and the resource-intensive nature of field monitoring, water managers need models that identify where and when lakes require additional monitoring. Models that rely on the spatial variation in cyanoHABs between lakes can help identify lakes at higher risk of developing cyanoHABs. While this approach lacks a

temporal component, narrowing the population of lakes requiring more attention is a useful analysis that can aid prioritization of resources among lakes.

Advances in remote sensing have increased the spatial and temporal extent of cyanoHABs data, providing information across the globe on an almost daily basis. Satellites cannot directly detect non-optically active parameters such as cyanotoxins, but can measure the phytoplankton biomass which is optically active (International Ocean Colour Coordinating Group, 2018). Optically active parameters include the photosynthetic pigments chlorophyll *a* in phytoplankton and phycocyanin in cyanobacteria. A number of satellite-based sensor platforms have been leveraged to examine phytoplankton in inland lakes (Ho et al., 2019; Naghdi et al., 2020; Shi et al., 2017; Wynne et al., 2008). Among these, there are several advantages of the Medium Resolution Imaging Spectrometer (MERIS) and the current Ocean and Land Color Instrument (OLCI) including (1) multispectral data can be used to distinguish cyanobacteria from other phytoplankton (Moradi, 2014; Palmer et al., 2015; Wynne et al., 2013a) and (2) the spatial (300 × 300 m) and temporal (daily-weekly) resolution of data is sufficient to capture a large number of inland waterbodies at a frequency more likely to detect short-lived blooms (Gómez et al., 2011; Lunetta et al., 2015; Matthews and Bernard, 2015; Sòria-Perpinyà et al., 2021). Despite the inability to detect toxins, satellite data has led to a vast increase in synoptic monitoring of cyanoHABs. Detecting cyanoHABs with satellite imagery has led to the ability to quantify spatial extent, temporal frequency, occurrence, and magnitude metrics for cyanobacteria in satellite monitored lakes (Clark et al., 2017; Coffer et al., 2020; Coffer et al., 2021; Mishra et al., 2019; Schaeffer et al., 2022; Urquhart et al., 2017). Remote sensing data on phytoplankton more broadly has also been used to explore the climate, watershed, and lake factors that explain the variability in blooms among lakes (Ho et al., 2019; Iiames et al., 2021; Marion et al., 2017; Song et al., 2021). Even with these substantial advances, there is still an acute need for managing public health by identifying which lakes are at higher risk for toxic cyanoHABs. The number of lakes captured through satellite data is an opportunity to address this need at broad spatial scales.

Essential to identifying cyanoHABs that are a risk to public health is the presence and concentration of cyanotoxins. A challenge has been that while satellite-based sensors can monitor cyanobacteria, the relationship between cyanobacterial abundance and toxin concentration is highly variable. The variability in the relationship between cyanobacteria biomass and toxin production depends on cyanobacteria community composition (Macário et al., 2015) and a suite of environmental factors (Davis et al., 2009; Davis et al., 2010). As a result, identifying sufficient signal between cyanobacteria and toxin concentration is difficult because of the factors that produce temporal variability in cyanobacteria toxin production. Therefore, developing models that relate cyanobacteria to toxin concentration spatially within a lake and over time is highly challenging (Stumpf et al., 2016a). In contrast, taking a spatial approach of comparing lake-level cyanobacteria to microcystin concentration for a population of lakes may mitigate some of the variability challenges. The availability of satellite imagery and national-scale field surveys is an opportunity to take a population-level approach to this issue. Unlike field monitoring for cyanoHABs, which focuses on lakes with known cyanoHAB issues, satellite and national surveys include both lakes regularly experiencing cyanoHABs as well as less affected or unaffected lakes.

As a result of the broad inclusion of lakes in these data sources, there is large range in the cyanobacteria-microcystin data compared to most within-lake sampling campaigns that target when cyanoHABs are occurring. This type of data has two related advantages. First, the large data range can help to characterize the relationship and constrain error. Second, the spatial variation between lakes likely exceeds the temporal variation within lakes. Spatial relationships require only that data be representative of the differences among locations. One test that assesses this representativeness is the signal-to-noise ratio (S/N) (Kaufmann et al., 1999), which compares the signal from some process to the background noise. Provided that the spatial variation in data among sites (signal) is larger than the temporal variation within sites (noise), then the data can be used to identify spatial patterns (Kaufmann et al., 1999). Given the large number of waterbodies used for drinking water sources and recreation, an approach identifying which lakes are at risk for toxic blooms can provide management-relevant information. Such an approach is a spatial prioritization rather than a temporal one, but this can help identify lakes that may need more resource intensive field monitoring.

This study develops an approach for combining satellite imagery and field survey data to assess the probability of toxic cyanoHABs among large lakes across the conterminous US. The approach models the probability of lakes experiencing elevated microcystin, cyanobacteria, and chlorophyll a among lakes during the summer recreational season rather than applying a temporal forecast approach that would identify when blooms will occur. This among-lake approach is innovative for two reasons: (1) The approach connects satellite-based estimates of cyanobacteria to cyanotoxins for a large number of lakes over a broad geographic region and (2) a spatial framework avoids the pitfalls of developing a temporal relationship between cyanobacteria and toxins. The approach is developed using data sources from the US but can be applied in other parts of the world where remote sensing and field data are available. The approach helps identify which lakes are at-risk for microcystin across broad geographic areas, thereby addressing a critical need to narrow the number of lakes that may need additional monitoring, community risk communication and education, and where more specific predictive model development can be prioritized (Rousso et al., 2020).

2 Materials and Methods

Models were constructed to estimate the probability of lakes exceeding demonstration thresholds for microcystin, cyanobacteria abundance, and chlorophyll *a*. While the study focus is risk of microcystin, cyanobacteria abundance and chlorophyll *a* are included to demonstrate the consistency of the spatial approach used here for identifying lakes at risk for potentially harmful blooms. The US National Lakes Assessments (NLA) conducted in 2007 and 2012 provided field data on microcystin, cyanobacteria abundance, and chlorophyll *a* (Sec 2.1). Satellite imagery from MERIS was used to determine lake-level summer bloom magnitude for cyanobacteria (Sec 2.2). These data sources were chosen for three reasons: (1) There is considerable overlap among the lakes captured in both datasets (~10% of lakes resolved by MERIS in the US were sampled in at least one of the two NLAs); (2) the NLA data is collected using consistent field and laboratory methods, thereby reducing error when comparing among sites; and (3) the MERIS imagery has the multispectral information

necessary to distinguish cyanobacteria from other phytoplankton at a spatial and temporal resolution appropriate for monitoring inland lakes. For these reasons, MERIS and the 2007 and 2012 NLA were the focus for this analysis; however, the contribution of this work lies in the approach rather than the specific data sources or geographic region used in this application of the approach. An overview of the workflow for the data preparation and analysis is shown in Figure 2.

2.1 Field CyanoHAB Data

Lake cyanobacteria conditions in the field were assessed using the US Environmental Protection Agency's (EPA) NLA program (Pollard et al., 2018). Data from the 2007 and 2012 NLA were used since these surveys bookend the MERIS data collection from 2008–2011. The NLA is part of the National Aquatic Resource Survey program that assesses the physical, chemical, and biological condition of waterbodies across the conterminous US on a five-year cycle. Each NLA selects lakes according to a probability-based survey design to be representative of the population of US lakes. The population is defined as lakes that are at least 4 ha in surface area for the 2007 NLA and 1 ha in 2012, in addition to having a depth of at least 1 m. The collection and analysis of microcystin, cyanobacteria abundance, and chlorophyll *a* are described in previous studies (Beaulieu et al., 2013; Loftin et al., 2016; Rigosi et al., 2014) and technical documentation (US EPA, 2007; US EPA, 2011; US EPA, 2012). Briefly, approximately 1000 lakes distributed across the country were visited once in June-September for each survey. Approximately 10% of lakes are resampled at a later date in the June-September window to evaluate temporal variability in the data collected. In addition, approximately a third of lakes included in the 2007 NLA were also included in the 2012 NLA. At each lake, a depth-integrated photic zone water sample is collected for microcystin, cyanobacteria abundance, and chlorophyll *a*. Samples are collected from the deepest point in the lake and generally exclude shoreline conditions. Microcystin and chlorophyll *a* samples were stored on ice and cyanobacteria abundance samples were preserved with Lugol's iodine solution. Chlorophyll *a* samples were filtered in the field immediately after collection. In the lab, microcystin samples were subject to three freeze-thaw cycles to lyse cells and release toxins into the dissolved phase prior to measurement by enzyme-linked immunosorbent assay. The detection limit of the microcystin analysis was 0.1 μg/L. Cyanobacteria abundances were identified and enumerated via microscope. Chlorophyll *a* samples were extracted with 90% acetone and analyzed by fluorometry.

2.2 Satellite Imagery of Lake Cyanobacteria

As a covariate of cyanoHABs risk, MERIS images were used to estimate lake cyanobacteria abundance throughout the summer. While there are other sources of satellite data, they lacked one or more advantages of the MERIS data. While OLCI on Sentinel 3 provides the necessary multispectral spectral data at the appropriate spatial and temporal resolution, there is insufficient nation-wide field data collected concurrently with this satellite mission. The Moderate Resolution Imaging Spectroradiometer (MODIS) has the spectral bands to separate cyanobacteria from other phytoplankton with daily coverage, but the spatial resolution of 1 km precludes use from many inland lakes. Land based satellites such as Landsat and Sentinel 2 provide higher spatial resolution but lack the necessary spectral bands or temporal revisit frequency needed. Hyperspectral missions can separate

cyanobacteria from other phytoplankton, but these platforms lack the broad spatial and temporal coverage needed for operational monitoring. For these reasons, MERIS was used for this analysis; however, the approach developed here can be applied to other data sources. For example, Sentinel 3 with field data from the 2017 and 2022 NLA, when available, is a future opportunity for application of the approach.

MERIS images were collected from 2008–2011 with a temporal resolution of approximately 2–3 days. MERIS imagery prior to 2008 was not included because data density was variable from 2002–2007 due to limited onboard recording and lack of direct broadcast reception over North America (Mishra et al., 2019). MERIS data collection was stopped in April 2012 prior to the start of the typical summer recreational season in North America (Mishra et al., 2019). MERIS imagery was used because this collection period is bookended by the 2007 and 2012 NLA. Pixel size is 300 by 300 m. The multispectral data collected by MERIS was processed using a spectral shape algorithm originally described in Wynne et al. (2008) to assess cyanobacteria abundance as the cyanobacteria index (CI). To remedy potential false positives from chlorophytes (Wynne et al., 2013b), Lunetta et al. (2015) updated the CI algorithm to add an additional exclusion criteria with full algorithm history detailed in Coffer et al. (2020). This algorithm ('CI_cyano' hereafter), is calculated based on two different regions of the spectral shape using the following equation:

$$SS(\lambda) = \rho_s(\lambda) - \rho_s(\lambda_-) + \{\rho_s(\lambda_-) - \rho_s(\lambda_+)\}\frac{(\lambda - \lambda_-)}{(\lambda_+ - \lambda_-)} \quad \text{(Eq. 1)}$$

where λ is the wavelength of the central band, and λ_+ and λ_- are the adjacent reference band wavelengths. The term $SS(\lambda)$ is the spectral shape at the central band wavelength and $\rho_s(\lambda)$ is the Rayleigh-corrected reflectance at the given wavelength. The CI_cyano is calculated at SS(681) where λ = 681 nm, λ_- = 665 nm, and λ_+ = 709 nm. A concave spectral shape—or a negative $SS(681)$—indicates cyanobacteria presence. For more specific identification of cyanobacteria, the same spectral shape algorithm is calculated with λ = 665 nm, λ_- = 620 nm, and λ_+ = 681 nm. When $SS(665)$ is greater than zero, this indicates that the presence of phycocyanin in cyanobacteria is depressing the reflectance of 620 nm and is further confirmation of the presence of cyanobacteria. CI_cyano has received extensive validation in the US against cyanobacteria cell counts (Lunetta et al., 2015), chlorophyll *a* (Seegers et al., 2021; Tomlinson et al., 2016), the presence of microcystin and cell counts (Mishra et al., 2021), and state-issued water advisories (Schaeffer et al., 2018; Whitman et al., 2022). In addition, the spectral shape algorithm has been validated for use in Hungary (Palmer et al., 2015) and the Caspian Sea (Moradi, 2014).

MERIS data were obtained by National Aeronautics and Space Administration (NASA) from European Space Agency through a data sharing agreement and were then processed by the NASA Ocean Biology Processing Group (https://oceancolor.gsfc.nasa.gov/projects/cyan/) using their standard satellite ocean color software package (l2gen; SeaWiFS Data Analysis System, SeaDAS, https://seadas.gsfc.nasa.gov) and a Shuttle Radar Topography Mission (SRTM)-derived 60-m land mask with updates to include missing lakes and reservoirs in Rhode Island and Massachusetts (Urquhart and Schaeffer, 2020). The SRTM land mask and SeaDAS processing are static in relation to waterbody size and did not

account for periods of drought or flood. Flags to indicate potential contamination due to cloud, cloud shadow, and adjacency effects from neighboring land pixels, and to identify snow- or ice-covered waterbodies, were also applied (Urquhart and Schaeffer, 2020; Wynne et al., 2008). Since mixed land-water and land-adjacent pixels are removed, the data are representative of the centers of lakes and exclude shorelines. Finally, lakes must have at least three reliably resolved pixels to be included. In total, 2,192 large lakes are resolved in the conterminous US.

The resolved lakes are distributed across 46 of the 48 states in the conterminous US (Urquhart and Schaeffer, 2020). Based on the lake area from the US National Hydrography Dataset, the lake sizes range from 1.25 km^2 to over 4,000 km^2 (Coffer et al., 2021). The resolved lakes are distributed across nine North American ecoregions, defined as areas with distinct compositions of biotic and abiotic factors including lithology, soil characteristics, vegetation, climate, land use, and hydrology that affect the ecosystem integrity (Omernik, 1995; Omernik and Griffith, 2014). Most resolvable lakes were located in the Eastern Temperate Forest (39.5%), Northern Forest (22.8%), and Great Plains (20.8%) ecoregions (Schaeffer et al., 2022). A smaller number of lakes were located within the Northwestern Forested Mountains (8.7%), North American Desert (4.7%), Mediterranean California (1.9%), Marine West Coast Forest (1.0%), and Tropical Wet Forest and Temperate Sierras (0.5%) ecoregions (Schaeffer et al., 2022). Of the resolved lakes, 13% are dammed (Lunetta et al., 2015).

The CI_cyano was summarized for each lake using the summer spatiotemporal mean of the bloom magnitude that is calculated with the following equation as (Mishra et al., 2019):

$$Area\ normalized\ bloom\ magnitude = \frac{\frac{1}{M}\sum_{m=1}^{M}\frac{1}{T}\sum_{t=1}^{T}\sum_{p=1}^{P} CI\_cyano_{p,t,m}}{Lake\ area\left(km^2\right)} \quad \text{(Eq. 2)}$$

where P (upper case) is the number of valid pixels in within a given image in the lake; T is the number of composite images in each month; M is the number of months over which the data were summarized; and p, t, and m (all lower case) are indexes of summation. Lake area was calculated as the number of pixels in the waterbody that could be resolved for cyanobacteria presence multiplied by the pixel area (0.09 km^2). The resulting number encapsulates all bloom conditions experienced by the lake normalized by the time period considered and lake area. The area normalized bloom magnitude was calculated for each lake for June through September from 2008 to 2011 and the annual estimates used to calculate the interannual mean for each lake. This interannual mean summer area-normalized bloom magnitude (“bloom magnitude” hereafter) characterizes the mean summer lake conditions over the four-year monitoring period.

2.3 Spatial Versus Temporal Variation in Data Sources

Since NLA samples are collected just once per summer and the two NLAs (2007, 2012) occurred in the summers preceding and following the satellite imagery (2008–2011), each field data component was evaluated for spatial and temporal variation (Herlihy and Sifneos, 2017; Stoddard et al., 2008). The S/N was used to compare the signal from a variable

to the background noise (Kaufmann et al., 1999). When spatial variation in data among sites (signal) is larger than the temporal variation within sites (noise), data collected at different time points can be used together to identify spatial patterns (Kaufmann et al., 1999). In national surveys such as the NLA, variables with high spatial variation relative to temporal variation confirm that between-site differences in individual samples are caused by differences in waterbody condition rather than by sample variation across time (Stoddard et al., 2008). The S/N is determined using a linear mixed effects model where the observed data value is the response variable (i.e., cyanobacteria abundance). Covariates in the model are the site variable (i.e., lake) as a random effect and the site visit number as a fixed effect. The signal is the variance associated with the site random effect in the model. The noise is the residual variation and is associated with variation over time and sampling error. The S/N is the ratio of the variance among all sites (signal) to the variance of the repeated visits to the same site (noise). S/N values >1 indicate that spatial variance is larger than temporal variance. S/N values ≤1 indicate that visiting a single site yields at least as much variability in the parameter of interest as visiting two different sites. Generally, for variables to be used as indicators of site characteristics, a S/N of >10 is considered robust (Kaufmann et al., 1999). A S/N of 2–6 is considered a moderately robust measure with some temporal variation. A S/N of <2 is considered a relatively poor measure of site condition because the temporal variability in the parameter is high relative to variability across sites. There are no sample size requirements for evaluating the S/N, only that the samples are sufficiently representative of the range of values in the metric. For this analysis, a S/N of >2 was used to indicate that comparing data collected in different years were appropriate for identifying spatial patterns in lake cyanoHABs risk.

The S/N was evaluated at two different temporal scales for NLA microcystin, cyanobacteria abundance, and chlorophyll *a*: (1) variation within the summer survey season using the revisit samples and (2) variation across the two NLAs using lakes included in both surveys. Within survey S/N was determined by comparing the variance among all sites sampled in each NLA to the variance within the ~10% of revisited sites. Across survey S/N was determined by comparing the spatial variance among all observations in the 2007 and 2012 NLAs to the temporal variance among the approximately one-third of lakes resurveyed. For cyanobacteria abundance and chlorophyll *a*, the respective cell density or concentration was used as the response variable. For microcystin, the analysis was run twice: first with only microcystin observations above the detection limit and second with microcystin observations below the detection limit recoded to have a value of zero. The S/N for the satellite-derived bloom magnitude was evaluated by comparing the bloom magnitude for each summer between 2008 and 2011 across the 2,192 satellite-resolved lakes.

As an additional illustration of spatial versus temporal variation in the satellite data, the variation in weekly bloom magnitude was compared to the variation among all satellite-resolved lakes for the interannual mean summer bloom magnitude. The weekly bloom magnitude was calculated as the sum of all pixel values divided by the lake area. The variation for each lake was computed using the variance among weekly bloom magnitude measurements in each lake collected between June and September for each year between 2008 and 2011. The mean across the four years of data collection for each lake was calculated. The variation among lakes was calculated as the variance in the interannual mean

summer bloom magnitude across the 2,192 satellite resolved lakes. The weekly variance (temporal variation) was compared to the among-lake variance (spatial variation) to illustrate how the within-lake temporal variation compares to the population-level spatial variation.

2.4 Compilation of Field Data for Comparison with Satellite Information

There were 210 lakes across 37 states that were both assessed for CI_cyano and sampled in one of the NLAs. This consisted of 149 lakes from the 2007 NLA and 122 lakes from the 2012 NLA. Among these were 61 lakes that were sampled in both the 2007 and 2012 NLAs (Fig 1). A dataset of independent observations was created that consisted of the 88 unique lake observations from the 2007 NLA, 61 unique lake observations from the 2012 NLA, and randomly selecting observations from either the 2007 or the 2012 NLA for the remaining 61 lakes that were sampled in both assessments. This process yielded a total of 210 distinct observations. By randomly selecting one observation from the resurveyed lakes, the assumption of independent observations is met. The effect of the random observation selection process and differences based on the year in which NLA data was collected was evaluated on the final models (see Supplemental Information).

2.5 Field Data Relationships

Relationships among the NLA microcystin, cyanobacteria abundance, and chlorophyll *a* were evaluated using a Spearman's rank correlation. Since >50% of microcystin observations were below the detection limit, the analysis was conducted with two different methods for these observations. First, only microcystin observations above the detection limit were included. Second, all microcystin observations were included by recoding all microcystin observations below the detection limit to have a value of zero. In the second case, the number of ties in the ranked data can result in an imprecise p-value estimation but presenting both sets of results is illustrative of the correlations between the cyanoHABs field variables with and without microcystin data below the detection limit.

2.6 Model Construction

The microcystin, cyanobacteria abundance, and chlorophyll *a* data were categorized based thresholds in order to estimate the probability of threshold exceedance based on the bloom magnitude. Two demonstration thresholds are included to show how the approach is flexible with respect to the chosen threshold level; however, thresholds need to be constrained to the range of values in the field data for the modeling to be successful. Since the NLA is designed to capture typical lake conditions rather than target cyanoHABs, the NLA cyanoHAB observations were mostly lower than measurements collected during active blooms. Therefore, for cyanobacteria and chlorophyll *a,* previously developed WHO lower (20,000 cells/mL and 10 μg/L) and higher (100,000 cells/mL and 50 μg/L) risk guidelines for recreational water use were used, respectively (World Health Organization, 2003). For microcystin, a lower threshold of 0.2 μg/L, a commonly reported detection limit for the toxin was used (Mishra et al., 2021), to indicate the probability of detecting microcystin. In addition, a higher microcystin threshold of 1 μg/L to demonstrate the utility of the approach for evaluating the probability of exceedance at a different, higher threshold level. For context, these microcystin thresholds are in the range of guidelines developed for finished drinking water exposure, which do not apply to the lake water measured in the NLA

but are below recreational guidelines (Chorus and Welker, 2021; US EPA, 2015; US EPA, 2019). With additional microcystin observations from active blooms extending the range of field values, the approach could be applied using higher threshold values that are closer to recreational guidelines for toxins during cyanoHAB events.

Logistic regression was used to model the odds of exceeding the lower and higher thresholds for microcystin, cyanobacteria abundance, and chlorophyll *a* with bloom magnitude as the sole covariate for a total of six models. Other in-lake environmental variables measured by the NLA were not included as additional covariates in the models because these data are only available for lakes that are both included in the NLA and are also satellite-resolved. Inclusion of this additional data would therefore limit model application to only those satellite-resolved lakes that were sampled in the NLA. The modeling results are presented for the 210 lakes as odds ratios, the factor by which the odds of threshold exceedance increase for a lake with 0.01 CI_cyano/km^2 greater bloom magnitude as compared to another lake. A unit of 0.01 CI_cyano/km^2 represents approximately 5% of the range of bloom magnitudes among the lakes used to develop the models.

The models were applied to all 2,192 satellite-resolved lakes to estimate the probability of exceeding each cyanoHAB threshold during summer. Based on the lake bloom magnitude, the estimated probability and 95% confidence interval of exceeding the cyanoHAB demonstration thresholds was retrieved for each lake using each of the six models.

2.7 Evaluating Model Robustness

Model performance was evaluated by conducting a leave-one-out cross validation (LOOCV) (Hastie et al., 2009). The leave-one-out validation approach was used rather than splitting the limited number of observations into calibration and validation datasets. For LOOCV, one observation is removed from the data and the model is generated using all other observations. The resulting model is used to estimate the probability of exceedance for the left-out observation. The process is repeated for each observation in the dataset. Finally, all estimated probabilities for the left-out observations are compared against their true values to determine model performance by area under the receiver operating characteristic (AUC), sensitivity (true positive rate), specificity (true negative rate), and model accuracy (true case identification rate)(Hosmer et al., 2013). The AUC quantifies the ability of the model to distinguish between classes for the response variable across all probability cutoffs. An AUC value of $0.7 \leq AUC < 0.8$ is considered acceptable, $0.8 \leq AUC < 0.9$ is considered excellent, and $AUC \geq 0.9$ is considered outstanding (Table S1)(Hosmer et al., 2013). Sensitivity is the true positive rate and is calculated as the number of true positive cases divided by the sum of true positive and false negative cases at a given probability cutoff (Table S2). Specificity is the true negative rate and is calculated as the number of true negative cases divided by the sum of true negative and false positive cases at a given probability cutoff. Finally, the model accuracy is true case identification rate and is calculated as the sum of true positive and true negative cases divided by the total number of cases. While AUC is evaluated across all probability thresholds, the accuracy, sensitivity, and specificity were evaluated at probability cutoffs that, for each of the six models, were relevant to the probability of getting a positive event.

3 Results

3.1 Correlations Among NLA cyanoHAB Metrics

Cyanobacteria abundance and chlorophyll *a* were significantly positively correlated in the NLA observations used to construct the models (N = 210, $\rho = 0.64$, $p < 0.001$; Fig S1). Microcystin was significantly positively related to cyanobacteria abundance when microcystin observations below the detection limit were excluded ($\rho = 0.27$, $p = 0.009$) and strength of the correlation increased when observations below the detection limit were included by assigning a value of zero ($\rho = 0.42$, $p < 0.001$). Microcystin was significantly positively related to chlorophyll *a* when microcystin observations below the detection limit were excluded ($\rho = 0.30$, $p = 0.003$) and strength of the correlation increased when observations below the detection limit were included by assigning a value of zero ($\rho = 0.51$, $p < 0.001$).

3.2 Spatial vs. Temporal Variation in Data Sources

In most cases, the spatial variation (signal) far exceeded the temporal variation (noise) for the NLA field cyanoHAB data (Table 1). Microcystin had consistently higher spatial variation than temporal variation (robust, S/N ≥ 10) regardless of the inclusion of observations below the detection limit. For chlorophyll *a*, the S/N for the 2007 NLA was moderate (S/N=2.7), the 2012 NLA was robust (S/N=12.0), and across the two NLAs was moderate (S/N=3.9). Cyanobacteria abundance was the only variable where temporal variation exceeded spatial variation with a poor S/N of <0.1 for the 2012 NLA and across the two NLAs. Cyanobacteria abundance had higher spatial variation than temporal variation in the 2007 NLA with a moderate S/N of 5.7.

The across summer S/N for the bloom magnitude as measured by satellite was moderate (S/N=3.6). When comparing the weekly variance in bloom magnitude for each lake (temporal variation) to the among lake population variance for interannual mean bloom magnitude (spatial variation), 84% of lakes have a weekly variance lower than the population-level variance in bloom magnitude (Fig. 3). In addition, variance in weekly bloom magnitude was positively correlated with the mean summer bloom magnitude (N = 2192, $\rho = 0.97$, $p < 0.001$).

3.3 Comparing Satellite and Field Data to Quantify Probability of CyanoHABs Across the US

The summer bloom magnitude derived from the satellite imagery was significantly and positively related to the probability of exceeding all tested thresholds for microcystin, cyanobacteria abundance, and chlorophyll *a* (Figs 4 and S2). In other words, the lake-level integrated summer bloom magnitude, which encapsulates the cyanobacteria conditions over resolvable lake area during the June-September period for 2008–2011, was positively related to one-time field measurements of cyanobacteria. For microcystin at the lower (0.2 μg/L) threshold, a 0.01 CI_cyano/km^2 increase in bloom magnitude was associated with an odds ratio of 1.42 (1.28–1.61 95% confidence interval) increase in the odds of threshold exceedance (Fig S2). The odds ratio—or change in the odds—is constant across the range of bloom magnitude values used to develop the models. For example, a lake with a bloom

magnitude of 0.05 CI_cyano/km^2 compared to a lake that has a bloom magnitude of 0.04 CI_cyano/km^2 had a 42% (28–61%) higher odds of exceeding the lower 0.2 μg/L threshold for microcystin. The interval of 0.01 CI_cyano/km^2 was chosen since this represents approximately 5% of the range of lake bloom magnitudes used for model development which was similar to the full range of bloom magnitudes among all satellite-resolved lakes (Fig S3). For microcystin at the higher (1.0 μg/L) threshold, a 0.01 CI_cyano/km^2 increase in bloom magnitude was associated with a 1.40 (1.26–1.59) fold increase in the odds of threshold exceedance. For cyanobacteria abundance at the lower (20,000 cells/mL) and higher (100,000 cells/mL) thresholds, a 0.01 CI_cyano/km^2 higher bloom magnitude was associated with a 1.24 (1.13–1.38) and 1.23 (1.11–1.37) fold increase in the odds of threshold exceedance, respectively. For chlorophyll *a* at the lower (10 μg/L) and higher (50 μg/L) thresholds, a 0.01 CI_cyano/km^2 higher bloom magnitude was associated with a 1.54 (1.33–1.82) and 1.43 (1.27–1.64) fold increase in the odds of threshold exceedance, respectively.

When applied to all 2,192 satellite-resolvable lakes in the conterminous US, the models can estimate the probability of exceeding field cyanoHAB thresholds (Fig 5). Across the six thresholds modeled, most lakes (74–96%) had less than a 50% probability of exceeding field cyanoHAB thresholds. The number of these lakes with a >75% probability of exceeding the lower thresholds for microcystin (0.1 μg/L), cyanobacteria (20,000 cells/mL), and chlorophyll *a* (10 μg/L) was 162, 122, and 335, respectively (Table 2). The number of lakes with a >75% probability of exceeding the higher thresholds for microcystin (1.0 μg/L), cyanobacteria (100,000 cells/mL), and chlorophyll *a* (50 μg/L) was 70, 29, and 63 lakes, respectively. Together, approximately 8–16% of the satellite resolvable lakes have >75% probability of exceeding the lower thresholds, and only 1–3% of lakes have >75% probability of exceeding the higher thresholds.

3.4 Model Performance

The random observation selection process for lakes that were included in NLAs had little effect on the final model coefficients. In addition, most models were similar regardless of the year in which the NLA data was collected (see Supplementary Information). Based on the leave-one-out cross validation, the performance for five out of six models was acceptable to excellent (AUC = 0.77–0.89; Table 3). Only the model for cyanobacteria at the 20,000 cell/mL threshold performed below this range with AUC = 0.68. Most models had higher sensitivity ranging from 0.73–0.95 than specificity ranging from 0.59–0.81, indicating the models generally perform better at identifying exceedances (true positives) than non-exceedances (true negatives). Thus, most models performed well overall and generally performed better at identifying exceedance events than non-exceedance.

4 Discussion

The approach developed here demonstrates how satellite imagery can be combined with national field surveys to identify which of the ~2,200 largest lakes in the US are at-risk for exceeding thresholds for the cyanotoxin microcystin, cyanobacteria abundance, and chlorophyll *a*. The relationship developed for microcystin is key because satellites cannot

directly monitor cyanotoxins though their presence is critical information for assessing risk to public health. In fact, the spatial approach used in this study is an innovation in the field for three reasons: (1) The lake-level risk estimates addresses many of the previously described challenges in relating satellite data to field cyanotoxins (Stumpf et al., 2016a); (2) the approach can be applied using data from other platforms and for different geographic areas; and (3) the risk estimates produced for each lake can be used for several endpoints including identifying which lakes may require more intensive monitoring and investigating the landscape variables associated with cyanoHABs.

4.1 Connecting Satellite Imagery to Field Cyanotoxins

CyanoHABs that produce cyanotoxins are the primary concern for water managers in assessing risk to the environment and public health. The process of modeling cyanotoxins based on satellite derived biomass is challenging for two reasons: (1) toxins are not optically active and cannot be directly monitored via remote imagery and (2) many factors lead to a variable relationship between the satellite measurement of changes in photosynthetic pigment fluorescence and toxin production. The challenges of relating satellite imagery to toxins is reviewed in detail by Stumpf et al. (2016a). Here, four of these challenges are summarized along with descriptions of how the approach developed in this study addresses these and in addition to being consistent with recommendations from Stumpf et al. (2016a).

First, there is a high degree of intra- and inter-annual variability in the relationship between cyanobacteria pigments and cyanotoxins such as microcystin (Greene et al., 2021; Stumpf et al., 2016a). As a result, a single, fixed relationship between these two variables should not be assumed for any waterbody. Relationships between photosynthetic pigments and microcystin may only be relevant for 1–2 weeks in any given waterbody (Stumpf et al., 2016a). Indeed, microcystin production varies based on the cyanobacteria community composition and a suite of environmental conditions (Davis et al., 2009; Davis et al., 2010; Macário et al., 2015). While this variability is highly consequential when attempting to relate imagery to cyanotoxins over time and space within a single waterbody, the analysis developed in this study does not rely on a fixed relationship between satellite-derived bloom magnitude and the risk of microcystin for each waterbody. Instead, there is a demonstrated consistent spatial relationship among two time- and space-integrated metrics (1) the summer bloom magnitude and (2) the risk of microcystin exceeding the modeled threshold. This is possible because both the MERIS imagery and NLAs include lakes spanning a wide range in bloom conditions from those that are heavily affected by cyanoHABs to those that are less or not affected by the phenomenon. The additional evidence from the S/N for these data sources suggests the variation in bloom magnitude and microcystin is due to differences between lakes rather than temporal variation within lakes. The wide range in the data combined with evidence that the data represent differences among lakes makes the approach developed here sound for identifying among-lake differences in microcystin risk.

Second, each 300 m by 300 m pixel from the MERIS imagery likely has large variation in cyanobacteria abundance and microcystin across that area. As a result, mapping of within-lake microcystin concentration is discouraged because these maps are likely to underrepresent the true variability in microcystin spatially within a lake. The method

here was developed in acknowledgment of the high degree of spatial variation in cyanobacteria and toxins within lake, and the method does not produce specific within-lake spatial information. Instead, the approach generates a lake-level probability of microcystin threshold exceedance based on a lake-level integrated measure of bloom magnitude.

Third, time-averaged estimates of cyanobacteria from satellite imagery are vulnerable to underestimation bias due to wind-driven mixing of the lake water column (Hu et al., 2010; Hunter et al., 2008; Wynne et al., 2010). Underestimating cyanobacteria abundance in the lake can propagate to the microcystin prediction. This concern is addressed by using weekly composite images that preserve the maximum value for each pixel. Preserving the maximum value helps increase the chances that the peak cyanobacteria presence under calm wind conditions will be captured (Stumpf et al., 2012). These composite images are used as the basis of the interannual mean summer bloom magnitude. By using composite images, the potential for underestimation bias due to mixing is mitigated and the bloom magnitude is therefore inclusive of periods with high cyanobacteria.

Fourth, a two-step modeling approach was promoted for predicting toxins from satellite imagery (Stumpf et al., 2016a). The first step is the algorithm that relates the satellite imagery to the field-based surrogate pigment measurement and the second step relates the field-based surrogate pigment measurement to the toxin concentration. This approach is useful where the goal is to predict toxin concentration both spatially and temporally within a given lake. The advantage of this two-step process is that each model component can be measured and updated independently (Stumpf et al., 2016a). This is especially important in the context of temporal predictions (i.e., forecasting) since the pigment-toxin relationship can change over very short periods of time. The approach developed here is unique in that the model does not produce spatial and temporal predictions within individual lakes. Instead, the method relies on a spatial relationship between lake-level summer bloom magnitude and the risk of exceeding microcystin thresholds during the summer period. The relationship between these two data sources is consistent over the period studied because the S/N test for the bloom magnitude and microcystin data suggests that each variable’s variation is more related to differences between lakes rather than temporal variation within lakes. Since the spatial relationship between lake-level satellite-derived bloom magnitude and microcystin risk is consistent across the population of satellite-resolved lakes and the models are not attempting to quantify risk over time within lakes, an independent model relating in-situ cyanobacteria abundance to microcystin concentration is not required.

There are four additional aspects of the approach develop here that are advantageous. First, the CI_cyano algorithm is a derivative algorithm that uses the shape of the absorbance spectra rather than the absolute value of absorbance. Derivative algorithms are advantageous because (1) they are less sensitive to errors in the atmospheric correction which become more of a concern with higher frequency data collection (Philpot, 1991) and (2) they can be applied to both water column cyanobacteria and scum-forming blooms (Matthews and Odermatt, 2015; Wynne et al., 2013a). Second, an advantage of using the NLA data is that consistent field and laboratory methods were used for microcystin collection and measurement. This minimizes error in comparing microcystin levels between waterbodies since there are analytical uncertainties in the measurement of microcystin (Qian et al.,

2015). Third, the analysis includes four orders of magnitude in the satellite data and two orders of magnitude in the microcystin data. Several orders of magnitude in data sources are needed to adequately describe the relationship with reasonable error approximations (Stumpf et al., 2016a). This requirement is met through inclusion of over 2,000 large lakes that span from those that regularly experience cyanoHABs to those less or unaffected by cyanoHABs. Fourth, it is common practice to log-transform cyanobacteria and microcystin data to achieve homoscedastic errors in a linear relationship; however, this practice obscures the large amount of error in these relationships (Stumpf et al., 2016a). The use of logistic regression in the approach developed here avoids this pitfall by not requiring a linear relationship between variables and thus not requiring log-transformation. Logistic regression constrains probability estimates and confidence intervals between 0 and 1; therefore, the errors in the models are reasonable proportions of probability estimates. Taken together, the data collection methods for variables used in the models, the use of a large population of lakes allowing for a large range in the data, and the modeling approach all contribute to the advantages of this analysis.

In summary, there are many previously identified challenges with quantifying relationships between satellite imagery and microcystin in context of generating predictions about the cyanotoxin concentration spatially and temporally within any given lake. The approach developed here relies on the large range in the cyanoHAB data across space that characterizes the across-lake relationship while constraining model error. This allows for addressing many of these challenges and create reasonable lake-level estimates of summer season probability of exceeding chosen thresholds for cyanoHABs. Though less informative about spatial and temporal dynamics within individual lakes, the approach developed here provides useful lake-level information with reasonable estimates of uncertainty for prioritizing lakes that may become a public health concern and require further monitoring.

4.2 Bloom Magnitude Relates to Cyanobacteria and Chlorophyll *a*

Similar to microcystin, the models demonstrate that bloom magnitude relates to the risk of cyanobacteria and chlorophyll *a* exceeding the demonstration thresholds. Previous studies have established through validation approaches that the CI_cyano relates field measurements of cyanobacteria and chlorophyll *a*. For example, the CI_cyano was validated against cyanobacteria cell counts in eight US midwestern and eastern states (Lunetta et al., 2015), microcystin detections and cell counts in 11 states (Mishra et al., 2021), and more than 1,500 state-issued cyanoHAB advisories or reported events across 22 states (Schaeffer et al., 2018; Whitman et al., 2022). In addition, the CI_cyano relates to field chlorophyll *a* measurements in 15 US states (Seegers et al., 2021; Tomlinson et al., 2016), the Caspian Sea (Moradi, 2014), and in Hungary (Palmer et al., 2015). The goal of these studies was validation and is distinct from the goal of the present study to produce a spatial approach for estimating risk among lakes; however, the present study similarly found that the CI_cyano (here summarized as the bloom magnitude) relates to field measurements of cyanobacteria abundance and chlorophyll *a* for 37 states across the US. That the CI_cyano consistently has a quantifiable relationship with field algal measurements as found in this study and validation studies is further evidence that the CI_cyano is a useful metric for cyanoHABs for a broad array of lakes.

4.3 Spatial Models of CyanoHABs Risk

Spatial models of cyanoHABs risk are useful because it can help identify which lakes may become a public health risk. Using the combined satellite data and field survey information, the models can be used to narrow from a population of more than 2,000 of the largest US lakes to less than 350 lake that have a high risk (>75%) of exceeding the lower thresholds for microcystin, cyanobacteria abundance, and chlorophyll *a*, and less than 100 lakes that have a high risk of exceeding the high thresholds for cyanoHABs. By focusing on the 1–16% of lakes with elevated risk of developing cyanoHABs, water body managers can use this information to identify which lakes are a higher potential risk to public health and can help prioritize which lakes will require further costly field and laboratory monitoring.

Another advantage to taking a spatial approach for identifying cyanoHAB risk in lakes is this allows for a landscape analysis of the factors that lead to cyanoHAB risk. There were regional clusters of lakes with higher probability of exceeding thresholds in the Midwest, Gulf coast, and Florida peninsula; however, even within these regions there was high variability in lake cyanoHAB risk. For example, in southern Minnesota and the coastal region of Louisiana, adjacent lakes can have high and low probability of exceeding cyanoHAB thresholds (Fig S4 and S5). The proximity of lakes with different probabilities indicates that while regional factors can play a role in lake cyanoHABs, individual waterbody and watershed factors appear to be influential as well (Ho and Michalak, 2019; Iiames et al., 2021; Taranu et al., 2012). Future efforts could build on the risk estimates generated by the approach developed here, supporting research to identify landscape, climate, and waterbody drivers of blooms, which could also allow expansion of these risk estimates to smaller lakes below the threshold for satellite monitoring.

4.4 Limitations

While this analysis provides an approach to estimate the risk of cyanoHABs for over two thousand geographically disparate lakes, there are several limitations to the models developed here. First, both the NLA and satellite data used for the analysis are from the summer. As a result, the models are most useful in lakes that are characterized by summer cyanoHABs. While this does limit the utility of this approach for lakes that experience blooms in other seasons (Paerl et al., 2001), recreational cyanoHAB exposure is more common in summertime. However, factors such as rising temperatures due to climate change may make cyanoHABs and recreational advisories more common in other seasons. A second limitation of the analysis is the data only capture the center-of-lake conditions and the resulting models estimate the probability of this lake area being above the given threshold. This may underrepresent the risk in lakes where cyanoHABs accumulate near shorelines due to wind advection (Huang et al., 2014). Thus, the model results should be interpreted as the risk of threshold exceedance near the center of the lake. The third limitation is that the results are from a subset of US lakes—specifically, larger lakes and reservoirs. This is a result of the pixel size of the MERIS sensor. Approaches that could extrapolate these modeling results to smaller lakes could be explored. The fourth limitation is that wind and wave action affect cyanobacteria detection via satellites by mixing the biomass vertically in the water column. The satellite sensors used in this study detect the near-infrared and red wavelengths of the electromagnetic spectrum near the surface of the lake. As a result,

whole lake cell density can be underestimated when a bloom is distributed throughout the water column, or biomass may be missed if it is partially or completely below the satellite's depth of detection. Lakes that experience blooms where cyanobacteria are regularly below the satellite's depth of detection could account for some of the false negatives produced by the models developed in this study. Finally, while the multi-spectral images collected by MERIS can distinguish cyanobacteria from other types of phytoplankton, the imagery cannot distinguish community composition. Therefore, the bloom magnitude only provides information about the total cyanobacterial community and cannot distinguish between, for example, toxic and non-toxic community members. The inability to distinguish between toxin and non-toxin producing species may in part account for the false positive rate for the microcystin models. However, new technological advances in hyperspectral satellites include the Deutsches Zentrum für Luft Earth Sensing Imaging Spectrometer (DESIS) on the International Space Station, the Italian Space Agency PRecursore IperSpettrale della Missione Applicativa (PRISMA), Germany's Environmental Mapping and Analysis Program (EnMAP), the European Space Agency Copernicus Hyperspectral Imaging Mission for the Environment (CHIME), and NASA's Surface Biology and Geology (SBG) that may prove useful in future analysis to distinguish community types.

4.5 Potential Applications

In addition to using satellite imagery for evaluating cyanoHAB risk among lakes, there is an acute need to predict cyanoHAB risk over time within lakes (Lunetta et al., 2015; US EPA, 2019). The present analysis is tested only for identifying probability among lakes across the US; however, applying this method to forecast cyanoHABs over time may be possible in lakes with the largest variation in bloom magnitude. Approximately 15% of the MERIS-resolved lakes have a variance in weekly bloom magnitude greater than the variance in mean summer bloom magnitude across all US MERIS lakes (Fig 3). For this subset of lakes that experience large weekly variation in bloom magnitude, the models developed here may prove useful in predicting threshold exceedance over time. Testing the model utility for forecasting will require concomitant field sampling and satellite data that is difficult to obtain over broad geographic areas (Topp et al., 2020). However, the NLAs conducted in 2017 and 2022 have concurrent satellite data and is a future opportunity with potential to predict cyanoHABs through time in US lakes.

A useful feature of the approach developed here is the flexibility in the modeled thresholds. The thresholds used in this analysis demonstrate the method and could be adjusted to reflect state action levels for issuing water advisories based on the waterbody use. For example, a waterbody that serves as a drinking water source may become a concern at a lower level of microcystin than a waterbody used solely for recreation. Thus, a lower threshold could be used to help identify which waterbodies in a given region may reach microcystin levels in raw water that are concerning near drinking water intakes. However, the modeled thresholds are limited by the range of values in the field data. For microcystin, few samples from the 2007 and 2012 NLA exceeded the US EPA recreational health guideline of 8 μg/L (Loftin et al., 2016; US EPA, 2016). Using the approach with higher threshold values will require the incorporation of additional field data that specifically targets sampling around cyanoHAB events, which would mean supplementing NLA data.

The present analysis is relevant to the period between the first two NLAs in 2007 and 2012 and to the conterminous US, but the approach can be applied using new data sources to other time periods and geographic regions. For example, the more recent OCLI data from Sentinel 3 combined with the 2017 and 2022 NLA survey data (when available) could be used to update the analysis. While this study is specific to the conterminous US, the MERIS and Sentinel 3 missions have global coverage. Presently, the data from these satellites are used for detecting and monitoring lake cyanoHABs in Europe (Gómez et al., 2011; Sòria-Perpinyà et al., 2021) and South Africa (Matthews and Bernard, 2015). Expected increases in resolution between 10–30 m with Landsat and Sentinel 2 satellite platforms would expand coverage of smaller waterbodies and increase the value of satellite imagery for assessing water quality. In addition, projects such as the Water Information System for Europe (WISE) are increasing large-scale field data collection and compilation (https://water.europa.eu/freshwater). In practice, the combination of satellites, *in-situ* sondes, field sampling, and modeling efforts provides the best available information for water quality risk assessments because these methods provide observations at different spatial and temporal scales. The multiple expanding data sources on lake cyanoHABs makes developing global relationships between satellites and field data a rapidly approaching future.

5 Conclusion

The approach developed in this study is a key advancement in the application of satellite-derived data and large-scale field surveys to evaluate cyanoHAB risk because it (1) connects to the risk of microcystin and (2) applies to more than 2,000 geographically disparate lakes across the US. The spatial approach to modeling the relationship between satellite-derived bloom magnitude and field cyanotoxins avoids many of the challenges and pitfalls that make a difficult task of temporally relating the optically active photosynthetic pigments in cyanobacteria to the non-optically active cyanotoxins. The large range in the cyanoHAB data from a broad sample of large lakes characterizes the across-lake relationship while constraining model error. Though less informative about spatial and temporal dynamics within individual lakes, the spatial approach focuses in on the lakes that have a high probability of exceeding thresholds for microcystin. With anticipated increases in the spatial and temporal resolution of satellite-based lake cyanoHABs data, applying the approach developed in this study to a broader population of lakes may be possible soon. Equipped with lake cyanoHAB risk information, water managers can (1) prioritize which lakes require additional resource-intensive field monitoring, (2) identify near-lake communities that may need targeted education about potential risks of cyanoHABs in drinking and recreational waters, and (3) evaluate for vulnerabilities in drinking water infrastructure that are downstream or rely directly on high-risk lakes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We would like to thank Michael Dumelle for assistance with statistical analysis and Karen Blocksom and Megan Coffer for assistance with data. This article has been reviewed by the Center for Public Health and Environmental

Assessment and approved for publication. The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

Funding:

This research was funded by the US Environmental Protection Agency.

Data and materials availability:

All data and code for this analysis will be available on Data.gov upon acceptance for publication.

References and Notes

Beaulieu M, Pick F, Gregory-Eaves I. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnology and Oceanography 2013; 58: 1736–1746.

Beaver JR, Manis EE, Loftin KA, Graham JL, Pollard AI, Mitchell RM. Land use patterns, ecoregion, and microcystin relationships in U.S. lakes and reservoirs: A preliminary evaluation. Harmful Algae 2014; 36: 57–62.

Brookfield AE, Hansen AT, Sullivan PL, Czuba JA, Kirk MF, Li L, et al. Predicting algal blooms: Are we overlooking groundwater? Science of The Total Environment 2021; 769: 144442. [PubMed: 33482544]

Chapra SC, Boehlert B, Fant C, Bierman VJ, Henderson J, Mills D, et al. Climate Change Impacts on Harmful Algal Blooms in US Freshwaters: A Screening-Level Assessment. Environmental Science & Technology 2017; 51: 8933–8943. [PubMed: 28650153]

Chorus I, Welker M. Toxic cyanobacteria in water: A guide to their public health consequences, monitoring and management. CRC Press, on behalf of the World Health Organization, Geneva, CH, Boca Raton, FL, 2021.

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. [PubMed: 30245589]

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. [PubMed: 34326705]

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.

Davis TW, Berry DL, Boyer GL, Gobler CJ. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. Harmful Algae 2009; 8: 715–725.

Davis TW, Harke MJ, Marcoval MA, Goleski J, Orano-Dawson C, Berry DL, et al. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of Microcystis during cyanobacterial blooms. Aquatic Microbial Ecology 2010; 61: 149–162.

Elliott JA. Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. Water Research 2012; 46: 1364–1371. [PubMed: 22244968]

Gómez JAD, Alonso CA, García AA. Remote sensing as a tool for monitoring water quality parameters for Mediterranean Lakes of European Union water framework directive (WFD) and as a system of surveillance of cyanobacterial harmful algae blooms (SCyanoHABs). Environmental Monitoring and Assessment 2011; 181: 317–334. [PubMed: 21243424]

Greene SBD, LeFevre GH, Markfort CD. Improving the spatial and temporal monitoring of cyanotoxins in Iowa lakes using a multiscale and multi-modal monitoring approach. Science of The Total Environment 2021; 760: 143327. [PubMed: 33239199]

Hastie T, Friedman J, Tibshirani R. The elements of statistical learning: data mining, inference, and prediction. Vol 2. New York: Springer, 2009.

Herlihy AT, Sifneos JC. Within-year temporal variability in water chemistry in EPA's National Aquatic Resouce Surveys (2000–2014). Oregon State University, Corvallis, OR, 2017.

Ho JC, Michalak AM. Exploring temperature and precipitation impacts on harmful algal blooms across continental US lakes. Limnology and Oceanography 2019; 9999: 18.

Ho JC, Michalak AM, Pahlevan N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. Nature 2019; 574: 667–+. [PubMed: 31610543]

Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. Hoboken, New Jersey, USA: John Wiley & Sons, 2013.

Hu C, Lee Z, Ma R, Yu K, Li D, Shang S. Moderate Resolution Imaging Spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. Journal of Geophysical Research: Oceans 2010; 115.

Huang JC, Gao JF, Hormann G, Fohrer N. Modeling the effects of environmental variables on short-term spatial changes in phytoplankton biomass in a large shallow lake, Lake Taihu. Environmental Earth Sciences 2014; 72: 3609–3621.

Hunter PD, Tyler AN, Willby NJ, Gilvear DJ. The spatial dynamics of vertical migration by Microcystis aeruginosa in a eutrophic shallow lake: A case study using high spatial resolution time-series airborne remote sensing. Limnology and Oceanography 2008; 53: 2391–2406.

Huo S, He Z, Ma C, Zhang H, Xi B, Xia X, et al. Stricter nutrient criteria are required to mitigate the impact of climate change on harmful cyanobacterial blooms. Journal of Hydrology 2019; 569: 698–704.

Iiames JS, Salls WB, Mehaffey MH, Nash MS, Christensen JR, Schaeffer BA. 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.

International Ocean Colour Coordinating Group. Earth Observations in Support of Global Water Quality Monitoring. In: Greb S, Dekker A, Binding C, editors. Reports and Monographs of the International Ocean Colour Coordinating Group. International Ocean Colour Coordinating Group, Dartmouth, Canada, 2018.

Kaufmann PR, Levine P, Robison EG, Seeliger C, Peck DV. Quantifying physical habitat in wadeable streams. US Environmental Protection Agency, Washinton, D.C., 1999.

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. [PubMed: 28073498]

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.

Macário IPE, Castro BB, Nunes MIS, Antunes SC, Pizarro C, Coelho C, et al. New insights towards the establishment of phycocyanin concentration thresholds considering species-specific variability of bloom-forming cyanobacteria. Hydrobiologia 2015; 757: 155–165.

Marion JW, Zhang F, Cutting D, Lee J. Associations between county-level land cover classes and cyanobacteria blooms in the United States. Ecological Engineering 2017; 108: 556–563.

Matthews MW, Bernard S. Eutrophication and cyanobacteria in South Africa's standing water bodies: A view from space. South African Journal of Science 2015; 111: 1–8.

Matthews MW, Odermatt D. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sensing of Environment 2015; 156: 374–382.

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. Science of The Total Environment 2021: 145462. [PubMed: 33609824]

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: 17. [PubMed: 30626902]

Moradi M. Comparison of the efficacy of MODIS and MERIS data for detecting cyanobacterial blooms in the southern Caspian Sea. Marine Pollution Bulletin 2014; 87: 311–322. [PubMed: 25148755]

Naghdi K, Moradi M, Rahimzadegan M, Kabiri K, Rowshan Tabari M. Quantitative modeling of cyanobacterial concentration using MODIS imagery in the Southern Caspian Sea. Journal of Great Lakes Research 2020; 46: 1251–1261.

O'Neil JM, Davis TW, Burford MA, Gobler CJ. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. Harmful Algae 2012; 14: 313–334.

Omernik JA. Ecoregions: a spatial framework for environmental management. In: David W, Simon TP, editors. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishing, Boco Raton, FL, 1995, pp. 49–62.

Omernik JM, Griffith GE. Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework. Environmental Management 2014; 54: 1249–1266. [PubMed: 25223620]

Paerl HW, S. Fulton R III, DybleI J, Moisander PH. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. The Scientific World 2001; 1: 76–113.

Palmer SCJ, Hunter PD, Lankester T, Hubbard S, Spyrakos E, N. Tyler A, et al. Validation of Envisat MERIS algorithms for chlorophyll retrieval in a large, turbid and optically-complex shallow lake. Remote Sensing of Environment 2015; 157: 158–169.

Philpot WD. The derivative ratio algorithm: avoiding atmospheric effects in remote sensing. IEEE Transactions on Geoscience and Remote Sensing 1991; 29: 350–357.

Pollard AI, Hampton SE, Leech DM. The Promise and Potential of Continental-Scale Limnology Using the U.S. Environmental Protection Agency's National Lakes Assessment. Limnology and Oceanography Bulletin 2018; 27: 36–41.

Qian SS, Chaffin JD, DuFour MR, Sherman JJ, Golnick PC, Collier CD, et al. Quantifying and Reducing Uncertainty in Estimated Microcystin Concentrations from the ELISA Method. Environmental Science & Technology 2015; 49: 14221–14229. [PubMed: 26516650]

Rigosi A, Carey CC, Ibelings BW, Brookes JD. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. Limnology and Oceanography 2014; 59: 99–114.

Rigosi A, Hanson P, Hamilton DP, Hipsey M, Rusak JA, Bois J, et al. Determining the probability of cyanobacterial blooms: the application of Bayesian networks in multiple lake systems. Ecological Applications 2015; 25: 186–199. [PubMed: 26255367]

Rousso BZ, Bertone E, Stewart R, Hamilton DP. A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. Water Research 2020; 182: 115959. [PubMed: 32531494]

Schaeffer BA, Bailey SW, Conmy RN, Galvin M, Ignatius AR, Johnston JM, et al. Mobile device application for monitoring cyanobacteria harmful algal blooms using Sentinel-3 satellite Ocean and Land Colour Instruments. Environmental Modelling & Software 2018; 109: 93–103. [PubMed: 31595145]

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.

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

Shi K, Zhang YL, Zhou YQ, Liu XH, Zhu GW, Qin BQ, et al. Long-term MODIS observations of cyanobacterial dynamics in Lake Taihu: Responses to nutrient enrichment and meteorological factors. Scientific Reports 2017; 7.

Song K, Fang C, Jacinthe P-A, Wen Z, Liu G, Xu X, et al. Climatic versus Anthropogenic Controls of Decadal Trends (1983–2017) in Algal Blooms in Lakes and Reservoirs across China. Environmental Science & Technology 2021.

Sòria-Perpinyà X, Vicente E, Urrego P, Pereira-Sandoval M, Tenjo C, Ruíz-Verdú A, et al. Validation of Water Quality Monitoring Algorithms for Sentinel-2 and Sentinel-3 in Mediterranean Inland Waters with In Situ Reflectance Data. Water 2021; 13: 686.

Stoddard JL, Herlihy AT, Peck DV, Hughes RM, Whittier TR, Tarquinio E. A process for creating multimetric indices for large-scale aquatic surveys. Journal of the North American Benthological Society 2008; 27: 878–891.

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. [PubMed: 28073474]

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. [PubMed: 22870327]

Taranu ZE, Zurawell RW, Pick F, Gregory-Eaves I. Predicting cyanobacterial dynamics in the face of global change: the importance of scale and environmental context. Global Change Biology 2012; 18: 3477–3490.

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.

Topp SN, Pavelsky TM, Jensen D, Simard M, Ross MRV. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. Water 2020; 12: 169.

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. [PubMed: 31871980]

Urquhart EA, Schaeffer BA, Stumpf RP, Loftin KA, Werdell PJ. A method for examining temporal changes in cyanobacterial harmful algal bloom spatial extent using satellite remote sensing. Harmful Algae 2017; 67: 144–152. [PubMed: 28755717]

US EPA. Survey of the Nation’s Lakes. Field Operations Manual, Washinton, DC, 2007.

US EPA. 2012 National Lakes Assessment. Field Operations Manual, Washington, DC, 2011.

US EPA. 2012 National Lakes Assessment. Laboratory Operations Manual. US Environmental Protection Agency, Washinton, DC, 2012.

US EPA. Drinking Water Health Advisory for the Cyanobacterial Toxin Microcystin, Washington, DC, 2015.

US EPA. National Lakes Assessment 2012: A collaborative survey of lakes in the United States. US Environmental Protection Agency, Washinton, D.C., 2016.

US EPA. Recommended human health recreational ambient water quality criteria or swimming advisories for microcystins and cylindrospermopsin. In: Water Oo, editor. US Environmental Protection Agency, Washinton, D.C., 2019.

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 U.S. lakes. Harmful Algae 2022; 115: 102191. [PubMed: 35623685]

World Health Organization. Guidelines for safe recreational water environments. Volume 1: Coastal and fresh waters. Geneva: World Health Organization, 2003.

Wynne TT, Stumpf RP, Briggs TO. Comparing MODIS and MERIS spectral shapes for cyanobacterial bloom detection. International Journal of Remote Sensing 2013a; 34: 6668–6678.

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, Fahnenstiel GL, Dyble J, Schwab DJ, et al. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: Development and initial evaluation. Journal of Great Lakes Research 2013b; 39: 90–99.

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.

Xin X, Zhang H, Lei P, Tang W, Yin W, Li J, et al. Algal blooms in the middle and lower Han River: Characteristics, early warning and prevention. Science of The Total Environment 2020; 706: 135293. [PubMed: 31846885]

Xu H, Qin B, Paerl HW, Peng K, Zhang Q, Zhu G, et al. Environmental controls of harmful cyanobacterial blooms in Chinese inland waters. Harmful Algae 2021; 110: 102127. [PubMed: 34887007]

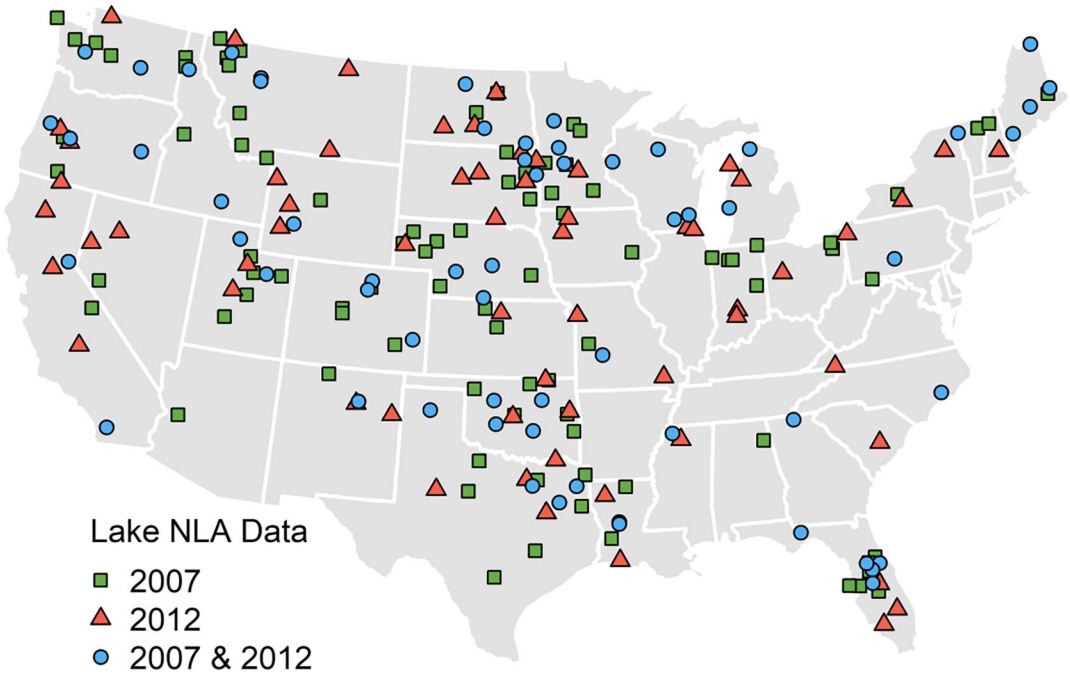

Fig. 1. Lakes with field and satellite data.

The location of the 210 lakes resolved by the MERIS sensor and sampled in either or both the 2007 and 2012 National Lake Assessments (NLA) for cyanobacteria harmful algal bloom metrics.

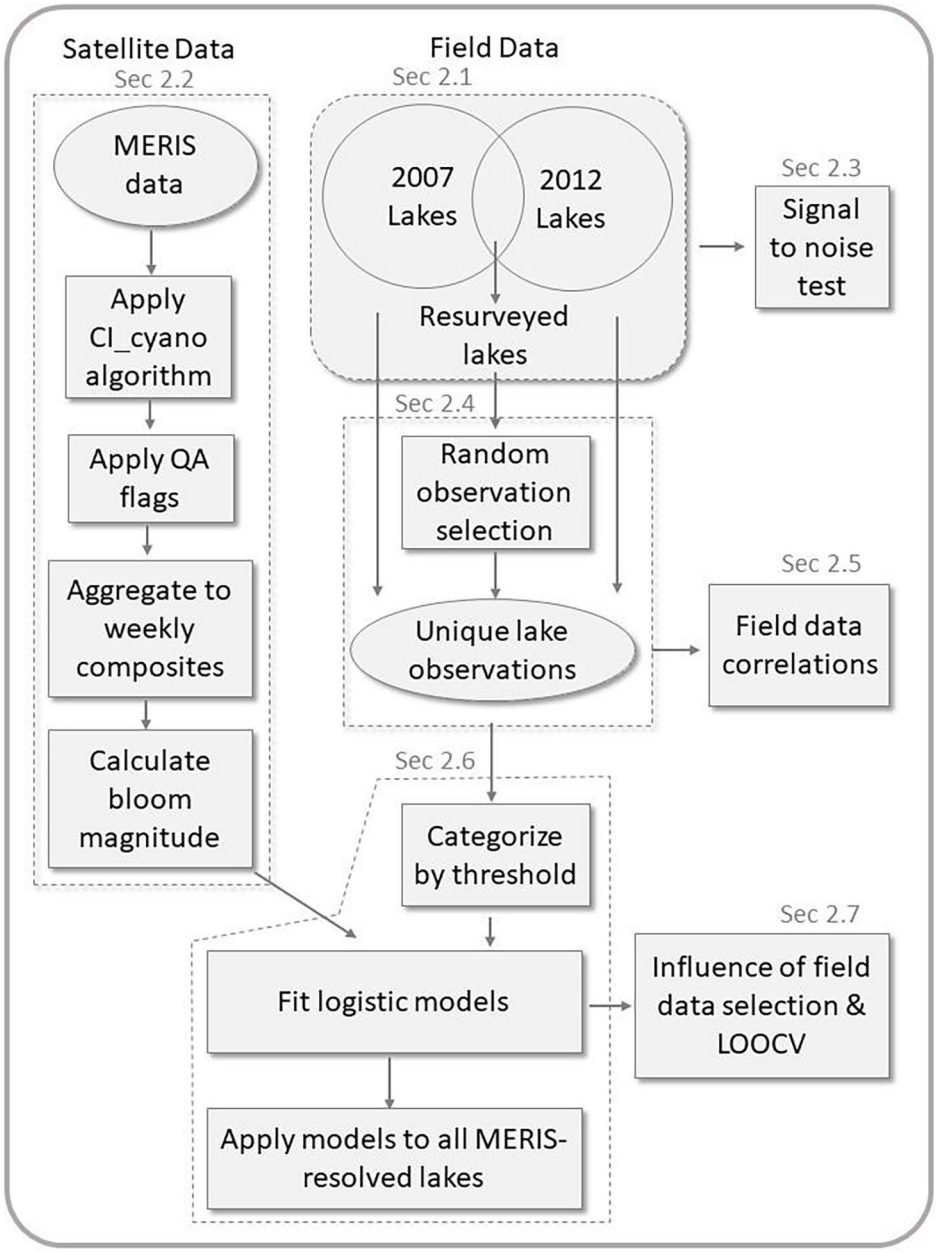

Fig. 2.
Schematic workflow diagram for analysis methods.

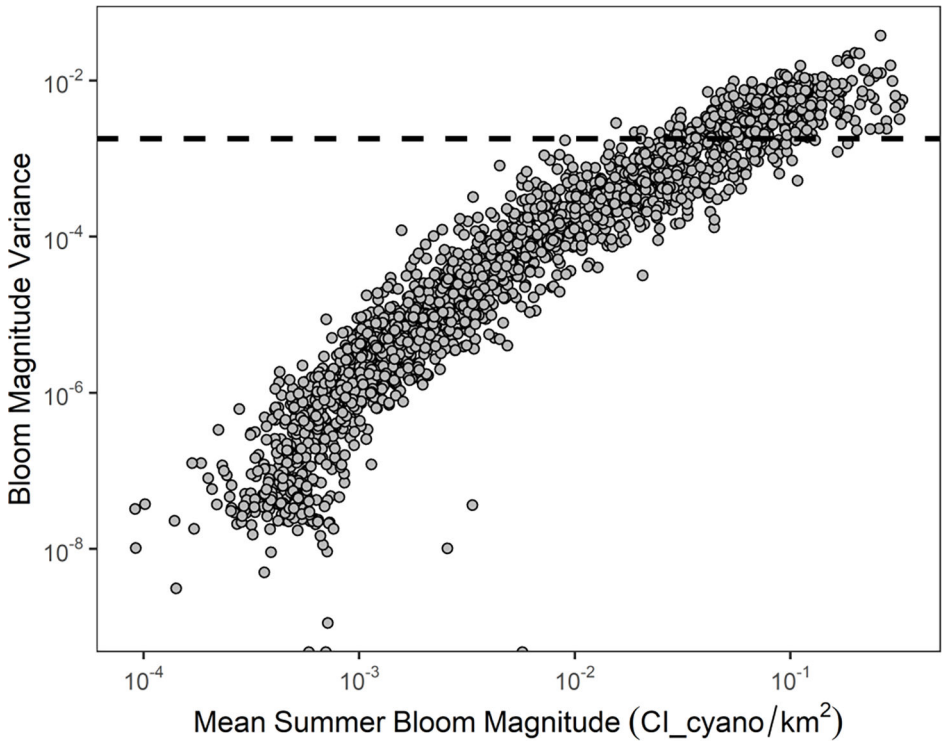

Fig. 3.

Variance in weekly area-normalized bloom magnitude versus mean summer bloom magnitude as measured by cyanobacteria index for individual satellite-resolved lakes (points, N = 2,192). Black dashed line is the variance in the interannual mean summer bloom magnitude among all satellite-resolved lakes.

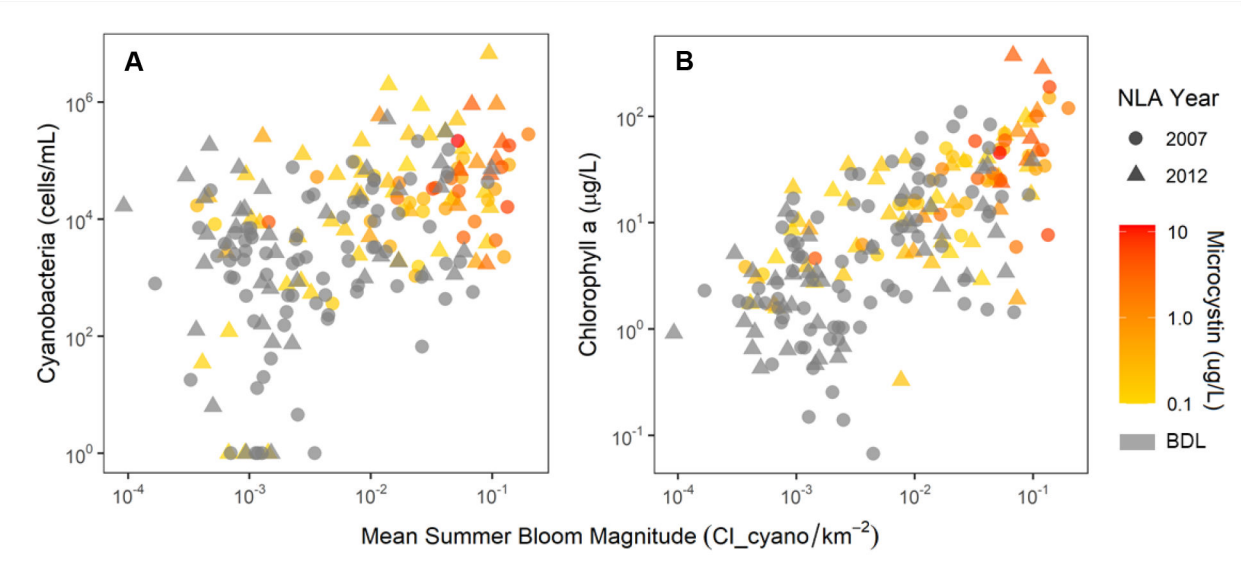

Fig. 4. Satellite and field data relationship.

National Lakes Assessment (NLA) cyanobacteria (**A**) and chlorophyll *a* (**B**) versus the interannual mean summer bloom magnitude measured by satellite. Warmer symbol colors represent high microcystin concentration and grey indicates observations below the detection limit (BDL, <0.1 µg/L).

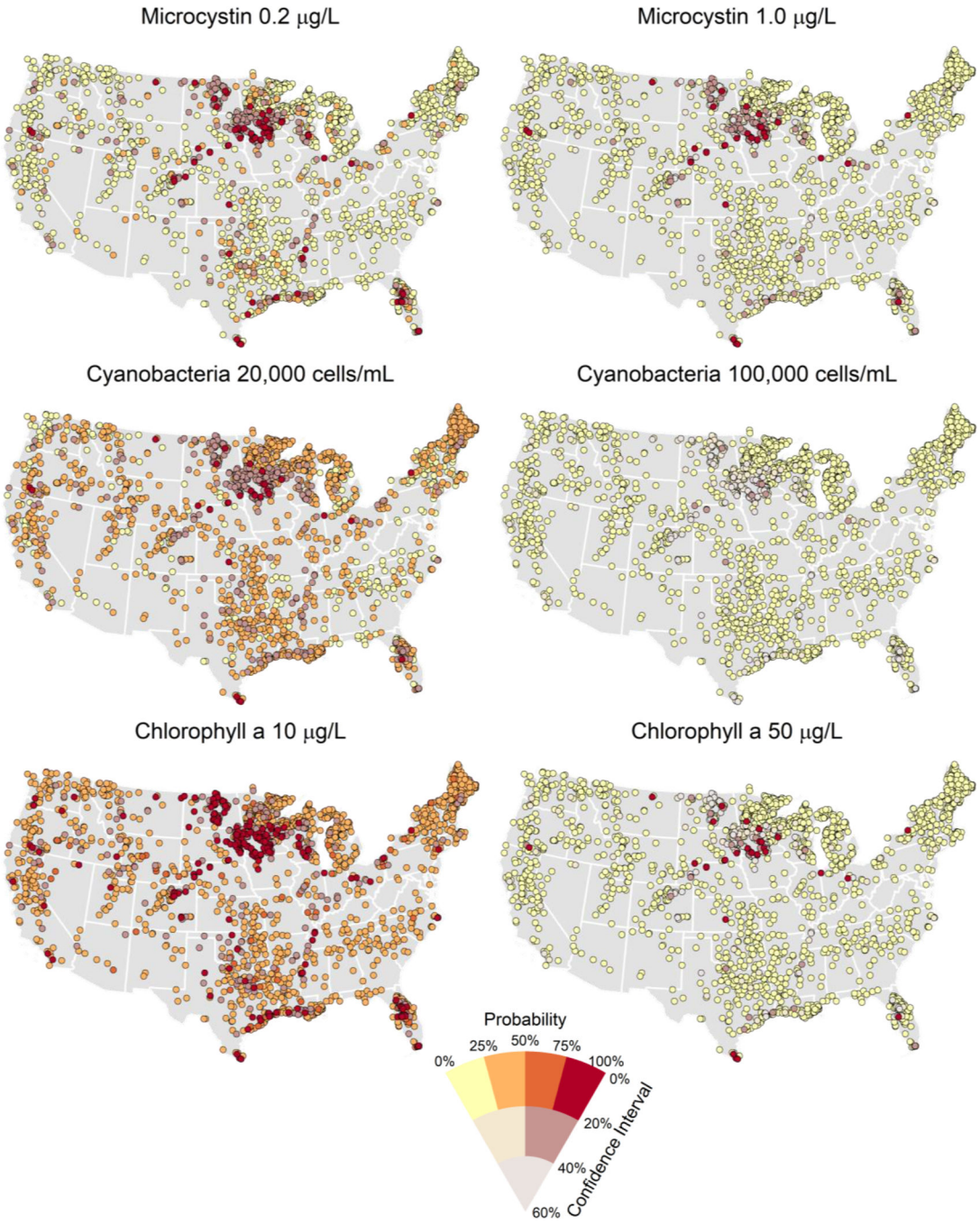

Fig. 5. Probability of cyanoHAB threshold exceedance.

Probability of exceeding lower and higher thresholds of microcystin, cyanobacteria, and chlorophyll *a* for all 2,192 lakes resolved by the MERIS sensor. Lower and higher thresholds are 0.2 and 1.0 µg/L microcystin (top row), 20,000 and 100,000 cyanobacteria cells/mL (middle row), and 10 and 50 µg/L chlorophyll *a* (bottom row). Warmer colors represent higher probability of exceedance. As the 95% confidence interval of the prediction widens, the symbol color desaturates and lightens.

Table 1.

The signal-to-noise (S/N) ratio for the National Lakes Assessment (NLA) cyanobacterial harmful algal bloom (cyanoHAB) metrics among all observations collected in the 2007 survey, all observations collected in the 2012 survey, and all observations pooled across the 2007 and the 2012 NLAs. To calculate S/N, the variation among lakes (signal) is divided by the variation among repeat samples collected in the same lake (noise).

Variable	S/N Within 2007 NLA	S/N Within 2012 NLA	S/N Across 2007 & 2012 NLAs
Microcystin[a]	10.0	15.3	10.5
Microcystin[b]	13.4	20.9	14.4
Cyanobacteria	5.7	<0.1	<0.1
Chlorophyll *a*	2.7	12.0	3.9

[a]Only microcystin concentrations above detection limit of 0.1 μg/L included

[b]Microcystin observations below the detection limit were included by recoding to have a value of zero

Table 2.

Number of lakes in probability quartiles for exceeding bloom thresholds.

The number (%) of lakes out of the 2,192 lakes resolved by the MERIS sensor that fall within probability quartiles by lower and higher thresholds for microcystin, cyanobacteria abundance, and chlorophyll *a*.

Probability	Microcystin	Cyanobacteria	Chlorophyll *a*
Lower Threshold	0.2 μg/L	20,000 cells/mL	10 μg/L
0 – ≤25%	1672 (76.3%)	586 (26.7%)	0 (0%)
25 – ≤50%	243 (11.1%)	1279 (58.3%)	1641 (74.9%)
50 – ≤75%	115 (5.2%)	205 (9.4%)	216 (9.9%)
75 – ≤100%	162 (7.4%)	122 (5.6%)	335 (15.3%)
Higher Threshold	1.0 μg/L	100,000 cells/mL	50 μg/L
0 – ≤25%	1937 (88.4%)	2008 (91.6%)	1979 (90.3%)
25 – ≤50%	113 (5.2%)	124 (5.7%)	95 (4.3%)
50 – ≤75%	72 (3.3%)	31 (1.4%)	55 (2.5%)
75 – ≤100%	70 (3.2%)	29 (1.3%)	63 (2.9%)

Table 3.

CyanoHABs risk model performance.

The performance of each model was measured via a leave-one-out cross validation and summarize by area under the curve (AUC), accuracy, sensitivity, and specificity.

Response Variable	Threshold	AUC	Accuracy	Sensitivity	Specificity
Microcystin	0.2 μg/L	0.77	0.74	0.77	0.73
Microcystin	1.0 μg/L	0.80	0.81	0.81	0.81
Cyanobacteria	20,000 cells/mL	0.68	0.66	0.79	0.59
Cyanobacteria	100,000 cells/mL	0.73	0.69	0.73	0.69
Chlorophyll a	10 μg/L	0.82	0.78	0.87	0.71
Chlorophyll a	50 μg/L	0.89	0.75	0.95	0.73