

Ground-based remote sensing provides alternative to satellites for monitoring cyanobacteria in small lakes

Katherine V. Cook^{a,b}, Jessica E. Beyer^{a,b}, Xiangming Xiao^d, K. David Hambright^{a,b,c,*}

^a Plankton Ecology and Limnology Laboratory, Department of Biology, University of Oklahoma, Norman, USA

^b Program in Ecology and Evolutionary Biology, Department of Biology, University of Oklahoma, Norman, USA

^c Geographical Ecology, Department of Biology, University of Oklahoma, Norman, USA

^d Center for Earth Observation and Modeling, Department of Microbiology and Plant Biology, University of Oklahoma, Norman, USA

ARTICLE INFO

Keywords:

Cyanobacteria
Remote sensing
Chlorophyll-*a*
Phycocyanin
Water quality
Harmful algal bloom
Monitoring

ABSTRACT

Cyanobacteria are the most prevalent bloom-forming harmful algae in freshwater systems around the world. Adequate sampling of affected systems is limited spatially, temporally, and fiscally. Remote sensing using space- or ground-based systems in large water bodies at spatial and temporal scales that are cost-prohibitive to standard water quality monitoring has proven to be useful in detecting and quantifying cyanobacterial harmful algal blooms. This study aimed to identify a regional 'universal' multispectral reflectance model that could be used for rapid, remote detection and quantification of cyanoHABs in small- to medium-sized productive reservoirs, such as those typical of Oklahoma, USA. We aimed to include these small waterbodies in our study as they are typically overlooked in larger, continental wide studies, yet are widely distributed and used for recreation and drinking water supply. We used Landsat satellite reflectance and *in-situ* pigment data spanning 16 years from 38 reservoirs in Oklahoma to construct empirical linear models for predicting concentrations of chlorophyll-*a* and phycocyanin, two key algal pigments commonly used for assessing total and cyanobacterial algal abundances, respectively. We also used ground-based hyperspectral reflectance and *in-situ* pigment data from seven reservoirs across five years in Oklahoma to build multispectral models predicting algal pigments from newly defined reflectance bands. Our Oklahoma-derived Landsat- and ground-based models outperformed established reflectance-pigment models on Oklahoma reservoirs. Importantly, our results demonstrate that ground-based multispectral models were far superior to Landsat-based models and the Cyanobacteria Index (CI) for detecting cyanoHABs in highly productive, small- to mid-sized reservoirs in Oklahoma, providing a valuable tool for water management and public health. While satellite-based remote sensing approaches have proven effective for relatively large systems, our novel results indicate that ground-based remote sensing may offer better cyanoHAB monitoring for small or highly dendritic turbid lakes, such as those throughout the southern Great Plains, and thus prove beneficial to efforts aimed at minimizing public health risks associated with cyanoHABs in supply and recreational waters.

1. Introduction

Cyanobacterial harmful algal blooms (cyanoHABs) in freshwaters are a topic of global concern (Paerl and Barnard, 2020). Blooms are increasing in frequency and magnitude due, in part, to increased anthropogenic nutrient loading (Paerl and Paul, 2012). CyanoHABs are particularly problematic because many species of harmful algae produce toxins that pose health risks to humans and other animals through exposure by toxin consumption, inhalation, and topical contact (Hambright et al., 2014; Hilborn and Beasley, 2015; Plaas and Paerl, 2021).

While some cyanotoxins may be only dermatotoxic, producing mild to moderate allergic reactions, common cyanotoxins, such as microcystin, cylindrospermopsin, saxitoxin, and anatoxin are powerful hepato- and neurotoxins that can cause chronic illness and death. With increasing threats of cyanoHAB development in inland waters, there is an equally increasing need for rigorous monitoring for cyanobacteria and their toxins to minimize public health risk associated with cyanoHABs (Paerl and Barnard, 2020).

Unfortunately, cyanoHAB monitoring today is highly insufficient for the needs of health risk management (Almuhtaram et al., 2021).

* Corresponding author.

E-mail address: dhambright@ou.edu (K.D. Hambright).

Historically, cyanoHAB assessment has been based on standard limnological monitoring of water quality with focus on criteria, such as water clarity or concentrations of chlorophyll-*a* and nutrients. Monitoring of cyanoHAB densities and cyanotoxins requires specialized skill and instrumentation and is thus generally conducted routinely only in lakes that provide water supply to large metropolitan communities. Similar routine cyanoHAB monitoring for the vast majority of lakes and reservoirs in rural settings is physically and economically impractical (Almuhitaram et al., 2021). High-frequency autonomous monitoring is possible with expensive, high-maintenance monitoring platforms (Coad et al., 2014). However, sufficient spatial coverage, particularly for moderate- to large-sized dendritic reservoirs, with tens to hundreds of semi-enclosed bays and coves (areas with high recreational use), would be prohibitively expensive. To minimize potential health risks associated with cyanoHAB toxin exposure in such systems, lake management and public health agencies need new tools that are amenable to broad and simple implementation across multiple and diverse systems, particularly those supporting high recreational tourism.

An early solution to the monitoring needs of public health protection emerged as remote measurement of solar reflectance by satellite-based sensors was shown to be effective and economically beneficial in assessing general chlorophyll-*a* concentrations in surface waters (Gons et al., 2002; Stroming et al., 2020). Satellite imagers have the potential to allow monitoring without the expensive person-hours and equipment required for a physical visit to a lake or a permanent autosampler/profiler, due to their large spatial coverage (~900–90,000 m²) which can collect images for many lakes simultaneously and their fine temporal resolution (daily to fortnightly flyovers, depending on satellite). Obviously, satellites are expensive, but many currently in use for earth observations were launched by public entities that offer users open-access data (images), thus providing lake managers a low-cost option for surface water observation.

Successful application of satellite-based remote sensing to cyanoHAB assessment, particularly for large waterbodies (Coffer et al., 2020; Urquhart et al., 2017; Wynne et al., 2010), has fueled a search for universal models that can predict concentrations of chlorophyll-*a* and phycocyanin, two key algal pigments commonly used for assessing total and cyanobacterial algal abundances (Shi et al., 2019). However, there are well-known limitations that may interfere with a satellite-based approach, including significant loss of usable images due to cloud cover (Ju and Roy, 2008) and sun-glint contamination (Overstreet and Legleiter, 2017). Moreover, satellite sensors frequently used for remote sensing of HABs (MODerate resolution Imaging Spectroradiometer, MODIS, aboard the National Aeronautics and Space Administration's (NASA) Terra and Aqua satellites; MEdium Resolution Imaging Spectrometer, MERIS, aboard the European Space Agency's (ESA) Envisat-1 satellite; and Ocean Land Colour Instrument, OLCI, a follow-up to ESA's MERIS) were built for a combination of atmosphere, terrestrial, and oceanic applications (Barnes et al., 1998; Nieke et al., 2012; Rast et al., 1999). The sensors on these satellites quantify reflectance from relatively large areas (e.g., 250 × 250 to 500 × 500 m) in each pixel. This resolution is not amenable for observing most inland lakes and reservoirs that are either too small or are highly dendritic in structure, such that many water pixels can be contaminated by littoral, shoreline, and terrestrial reflectance (Verpoorter et al., 2014; Wetzel, 2001). Additionally, water levels in smaller water bodies fluctuate seasonally, making the delineation between water and non-water contamination notoriously difficult (Zou et al., 2017), especially with low spatial resolution instruments. These spatial resolution limitations have led to a lack of data for smaller waterbodies in the remote sensing literature and the call for development and use of satellites with spatial resolutions of 30 × 30 m or smaller (Beck et al., 2016; Urquhart et al., 2017; Coffer et al., 2020).

An alternative to satellite-based remote sensing is the use of ground-based sensors, which do not suffer from the previously mentioned satellite limitations, like atmospheric interference (Wu et al., 2019).

Ground-based instruments with multispectral sensors have many advantages to satellites, like the ability to capture images with very fine spatial resolution, flexible temporal resolution, and rapid data turn-around time; studies suggest ground-based instruments will provide better capability for monitoring cyanoHABs than satellite-based remote sensing (Wu et al., 2019). There are commercially available sensors for use on ground-based instruments, but most were designed for terrestrial application. No commercially available sensors have a band designed specifically for phycocyanin. Fernandez-Figueroa et al. (2021) found commercially available cameras and multispectral sensors were more sensitive to chlorophyll-*a* than phycocyanin in eutrophic ponds. Multispectral sensors built for detection of phycocyanin are needed for adequate monitoring of cyanoHABs (Almuhitaram et al., 2021; Fernandez-Figueroa et al., 2021).

In this study, we sought a universal multispectral reflectance model that could be used for rapid, remote detection and quantification of cyanoHABs in reservoirs that are relatively small or highly dendritic, productive, and often turbid, like those in Oklahoma. We aimed to explore the effectiveness of satellite-based multispectral models for predicting chlorophyll-*a* and phycocyanin concentrations in Oklahoman reservoirs in comparison to new ground-based multispectral models derived from the same Oklahoma systems. Our results show that the published models had little to no predictive power with respect to assessing cyanoHABs in a study system outside those used to establish the models, and that ground-based sensors and multispectral models, particularly for phycocyanin, were far superior to Landsat-based sensors and models for assessing cyanoHABs in small, dendritic, productive reservoirs in Oklahoma.

2. Materials and methods

2.1. Study sites

Oklahoma is an ideal study system for turbid, small- to medium-sized reservoirs. The majority of Oklahoma lakes are smaller than the average American football field (i.e., ca. 0.5 ha, Zou et al., 2017) or are highly dendritic, with high shoreline to surface area ratios (aka, shoreline development; (Wetzel, 2001)). Oklahoma lakes are typical of lakes in the southern Great Plains with high turbidity, high chlorophyll-*a*, and frequent cyanoHABs (Oklahoma Water Resources Board, 2017). Lakes in Oklahoma are used primarily for water supply and recreation, and therefore monitoring harmful algal blooms is necessary to protect public health (Smithee et al., 2012).

With these characteristics in mind, we compared reflectance-algal pigment models using Landsat satellite data and ground-based hyperspectral reflectance. We built empirical models relating satellite spectral reflectance to chlorophyll-*a* (Table. S1) and phycocyanin (Table. S2) concentrations from 38 lakes (*N* = 1060 paired observations) sampled between 2001 – 2017 and ground-based reflectance to chlorophyll-*a* (*N* = 124) and phycocyanin (*N* = 125) concentrations from 7 lakes sampled between 2012 – 2016 (see Fig. 1A and B for map of sample locations). We also compared our Oklahoma-specific models to published models.

2.2. Satellite reflectance data

The Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI) are multispectral sensors onboard the Landsat 7 and Landsat 8 satellites, respectively, that consist of different spectral bands (ETM+ has eight bands and OLI has 11 bands) that record select areas of the electromagnetic spectrum (please see Irons et al. (2012) for details on the sensors and bands). These two satellites were chosen over other potential sensors (e.g., MERIS, OLCI, MSI) for this study due to the relatively small pixel size (30 × 30 m) of the ETM+ and OLI sensors and the eight-day revisit time between the satellites. Sentinel-3A OLCI was not chosen due to its large pixel size of 300 m which is difficult to resolve on small, dendritic Oklahoma reservoirs (see Fig. S6 for Landsat and

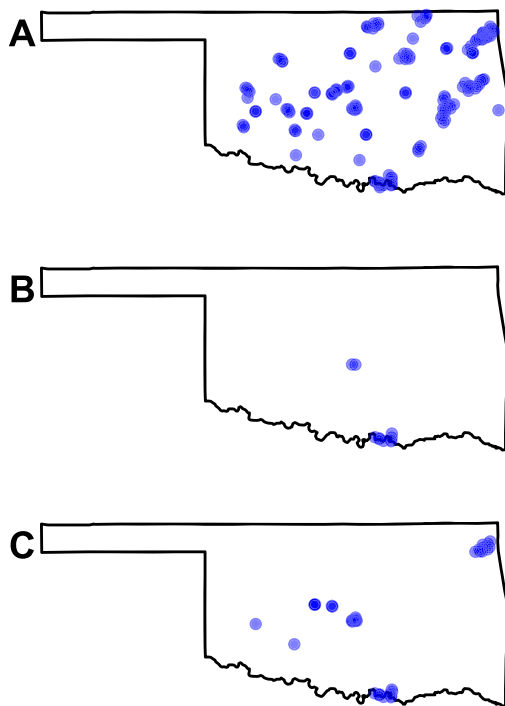


Fig. 1. Map of study sites in the US state of Oklahoma for (A) satellite-based chlorophyll-*a* model ($N = 191$), (B) satellite-based phycocyanin model ($N = 11$), and (C) ground-based multispectral chlorophyll-*a* and phycocyanin models ($N = 45$). Each point represents one site. Darker colors indicate overlap of geographically close sites.

Sentinel-3A image comparison) and it only became operational in 2016, years after much of our *in-situ* water sampling. Sentinel-2A MSI has a smaller pixel size (10 and 20 m) to Landsat (30 m) and a few more spectral bands that might be more useful than Landsat. One limitation of Sentinel-2A MSI, much like Landsat, is that it does not have a spectral band to measure phycocyanin. In addition, it was launched in 2015, after which only ~7% of our *in-situ* samples occurred. Thus, use of Sentinel-2A MSI would markedly decrease the data available for modeling in this study.

Surface reflectance data were obtained from EROS Data Center, Sioux Falls, SD (USGS) for each site-date combination. Surface reflectance data are generated by the USGS from Level 1 Digital Number (DN) products by applying a MODIS/6S atmospheric correction routine using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) software (Masek et al., 2006). Images ($N = 247$) were matched with *in-situ* measurements within a ± 7 -day window of the sampling date, ensuring a difference of no more than seven days between the satellite image and the *in-situ* measurements. This window was informed by a concurrent study measuring temporal autocorrelation in cyanobacterial blooms in Oklahoma (Beyer and Hambright, 2017) which showed that proxies (chlorophyll-*a* and phycocyanin) for cyanobacterial abundance during bloom periods were strongly temporally autocorrelated over a period of 10–11 days.

Images were visually screened for cloud cover and excluded if the lake area had >10% cloud cover. Since July 14, 2003, the Landsat 7 ETM+ Scan Line Corrector has been turned off due to an instrument failure, leaving stripes of missing data across images (Markham et al., 2004). To minimize the potential impact of missing data, and to account for the anchored swing radius of the boat under changing winds and currents, we averaged reflectance values from nine pixels (a 3×3 grid consisting of the sampling point pixel and the eight surrounding pixels) for each site. This approach maximized the number of usable data points and achieved the most representative sample for each reflectance measure (Baban, 1993). Reflectance data were extracted from images using

R (R Development Core Team, 2014) (see Supplemental for code). The data were then plotted and screened for high reflectance outliers (reflectance greater than 1). Water reflectance is much lower than land reflectance, so high reflectance values could indicate shoreline, or bottom interference, or an object (e.g., a boat) in the pixel.

2.3. Ground-based reflectance data

We collected *in-situ* hyperspectral reflectance data from seven lakes (see Fig. 1C for map) over a period spanning 2012–2016 using an ASD FieldSpec spectroradiometer (Malvern Panalytical, Malvern, United Kingdom). At each lake, the ASD was calibrated using a Spectralon reflectance panel (Labsphere, North Sutton, NH, USA). At each pelagic site ($n = 14$) and shoreline site on Lake Thunderbird ($n = 12$), the optical probe was held over the non-shaded side of the boat or boat dock at a 45-degree angle to the water and five spectra were recorded. These five spectra were averaged for each site to minimize the impact of sun glint or water waves and the resulting spectra were used in downstream analyses. For many of these lakes, we had repeat visits to multiple sites (Table. S3) and where feasible we took measurements at multiple locations on the lake to increase our spatial coverage. In total, we collected hyperspectral data from 132 site-date combinations.

Quality control and quality assurance were performed on the spectra. Spectra were visually examined, and six spectra were discarded: three due to the loss of the raw files, one due to an abnormal horizontal spike at 1000 nm, one due to abnormal low reflectance values, and another due to an outlier in the limnological data. Each spectrum was matched with same day *in-situ* pigment data. One additional site-date combination was discarded due to lack of corresponding limnological data and another was missing only chlorophyll-*a* data. The final working sample size was 125 for phycocyanin and 124 for chlorophyll-*a*.

2.4. In-situ surface water pigment data

Each satellite image and ground-based measurement has a corresponding pigment measurement from an Oklahoma lake. These data were collected by personnel in the Plankton Ecology and Limnology Lab (PELL) at the University of Oklahoma, as well as the Oklahoma Water Resources Board (OWRB) and the Grand River Dam Authority (GRDA). For the satellite dataset, 38 lakes were sampled between 2001–2017 for a total of 1060 chlorophyll-*a* data points (Fig. S1; Table. S1) and 97 phycocyanin data points (Fig. S3; Table. 2). For the ground-based measurements, seven lakes were sampled between 2012–2016 for a total of 124 chlorophyll-*a* (Fig. S4; Table. S3) and 125 phycocyanin data points (Fig. S5; Table. S3).

Lake sampling by PELL is briefly summarized here, for details see Hambright et al. (2010). At each site, a Hydrolab DS5X (OTT Hydromet, Kempen, Germany) sonde was used to measure phycocyanin (PCY) concentrations (starting in 2009). Although PCY in nature will generally be bound within cells, the Hydrolab PCY sensor was calibrated using serial dilutions of biologically relevant levels of pure PCY solution (Sigma Aldrich) and thus provides an underestimate of the actual amount of PCY in lakes. Nevertheless, relative changes can still provide meaningful insight into seasonality and site-to-site comparisons. Such PCY measurements have previously been found to correlate with cyanobacterial biomass (Thomson-Laing et al., 2020). At deeper sites, measurements were taken in the upper 10 m of the water column at 1-m intervals, and at shallow sites measurements were taken at 0.5-meter intervals through the entire water column (on average around 2 m). Phycocyanin measurements were averaged across the photic zone (as determined by a LI-COR light meter (LI-COR Biosciences, Lincoln, NE, USA) or 2.5 times the Secchi depth). For acetone-extracted chlorophyll-*a*, depth-integrated water samples were taken from the photic zone depth or the entire water column, whichever was shallower, and placed into acid-washed, deionized- and sample-rinsed Nalgene bottles, and kept on ice prior to filtration. The water was filtered onto

25-mm-diameter GF/F filters and stored in the freezer until chlorophyll-*a* extraction. In a reduced light environment, the filters were homogenized with a small volume of 70% acetone using a mortar and pestle, then incubated in 70% acetone at 4–8°C for 6–24 h. Samples were centrifuged prior to analysis on a Turner Designs 700 (pre-2012) or Trilogy (2012–present) bench-top fluorometers (Arar and Collins, 1997).

The OWRB and GRDA collected water samples from the upper 0.5 m for chlorophyll analysis. The agencies did not measure phycocyanin and both used the previously described EPA chlorophyll-*a* acetone extraction method. For the satellite dataset, OWRB contributed 836 data points, GRDA contributed 56, and PELL contributed a total of 168. For the ground-based dataset, all chlorophyll-*a* ($N = 124$) and phycocyanin ($N = 125$) pigment data points were contributed by PELL.

2.5. Landsat multispectral models

To quantify cyanoHABs in Oklahoma, we built empirical linear regression models using the chlorophyll-*a*, phycocyanin, and satellite reflectance values (see Fig. 2 for methods diagram). We used Landsat band ratios as our predictors because spectral ratios have been shown to be more robust than single bands (Vincent et al., 2004). We eliminated potential collinearity in the models using variance inflation factors (VIF) by sequentially removing the predictors with the highest VIF score until all scores were at the predetermined threshold of 10 (Zuur et al., 2010). The spectral ratios used to determine the final chlorophyll-*a* model included the following band ratios: Blue:Green, Red:Blue, Green:Red, Blue:Near Infrared (NIR), and NIR:Red; and in the PCY model: Blue:Green, Green:Red, NIR:Red, and Red:NIR. We used Bayesian Information Criterion (BIC) to select the model that explained the most variance in the response variable using the fewest predictors (Schwarz, 1978). All statistical analyses were completed in the R environment (version 4.2.1).

To test how published models effectively predicted cyanoHABs in Oklahoma lakes, we applied a set of chlorophyll-*a* algorithms adapted to Landsat 7 from the Landsat 8 models (Landsat 8 bands 2–6 match Landsat 7 bands 1–5) in Beck et al. (2016) (Table 1). The algorithms from the literature include the Normalized Difference Chlorophyll Index (NDCI) (Mishra and Mishra, 2012), the Surface Algal Bloom Index (SABI) (Alawadi, 2010), the Fluorescence Line Height algorithm focusing on the blue band (FLH blue) (Zhao et al., 2010), the two-band algorithm (2BDA) (Gitelson et al., 2003), the three-band algorithm (3BDA) (Dall’Omo and Gitelson, 2005) and the three band-like algorithm (KIVU) (Brivio et al., 2001; Kneubühler et al., 2007). We also used one established phycocyanin Landsat algorithm from the literature (Vincent et al., 2004) to compare with our Landsat phycocyanin model (Table 1). These algorithms were chosen in part because they were built

using Landsat bands or because they have been shown to successfully predict chlorophyll-*a* or phycocyanin.

Each of the previously mentioned chlorophyll-*a* models was applied to our Oklahoma satellite reflectance data set ($N = 1060$) and compared to *in-situ* chlorophyll-*a* values to determine model accuracy. As per the original use, each algorithm was solved prior to regressing (meaning there is a single regression coefficient for each algorithm). For the Vincent et al. (2004) phycocyanin algorithm, we fit a regression coefficient to each spectral ratio as was done in the original model. To assess the fit of the regressions, we compared the *in-situ* observed pigment values with the predicted algorithm pigment values and evaluated the relationship using the root mean square error (RMSE; lower values indicate better model fit) and Pearson’s r correlation test (range from +1 to –1 with numbers closer to 0 indicating no correlation) (Table 3). We also report the adjusted R^2 of each regression as this is a direct measure of model predictability (ranges from 0 to 1 with higher values being more predictive).

In addition to the above-described empirical models, we built random forest models to explore the utility of machine learning as an approach for generating useful predictive models. We used the randomForest function in the randomForest package in R (Liaw and Wiener, 2002) to grow random forests with $n_{tree} = 500$. For both chlorophyll-*a* and phycocyanin models we used the non-collinear predictors described above for the linear models. We chose regression random forest because we are most interested in prediction across a range of pigment values and they are more comparable to previously described linear models. We evaluated model performance similarly to the linear models using RMSE and R^2 but in addition, we tested significance against a null model. We extracted the performance statistics and significance using the rfUtilities package in R using the rf.regression and rf.significance functions, respectively (<https://cran.r-project.org/web/packages/utility/index.html>).

2.6. Ground-based multispectral models

We built a semi-empirical model using the ground-based measurements by first selecting five bands based on spectral properties of optical constituents of cyanobacteria and other non-target components that could affect reflectance of light from Oklahoma reservoirs, specifically, chlorophyll-*a*, phycocyanin, and turbidity (see Fig. 3 for methods diagram). The medians of the bands were based on published values (Matthews, 2011) and the width of the bands were optimized by comparing correlations of candidate bands with our measured optical constituents (Table 2). We calculated the average reflectance over the wavelengths included in the band. Using the bands (described in Table 2), we constructed candidate linear models for chlorophyll-*a*

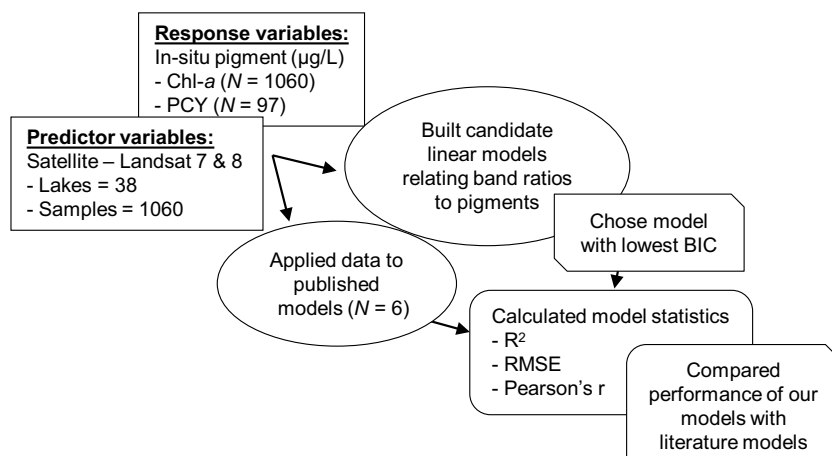


Fig. 2. Diagram showing satellite model development methods. See methods for details.

Table 1

Band math for each algorithm (derived from Beck et al. 2016, Vincent et al., 2004, and Randolph et al., 2008) used to estimate chlorophyll-*a* (1–5, 7, 9–10) and phycocyanin (6, 8–10) concentrations in Oklahoma lakes and each original citation. In addition, we have listed the type of algorithm and the original location and the algorithm for the relevant sensor. Ground-truthed data refers to pigment concentrations from waterbodies that were used to calibrate (build the model) or validate (test the model). Please see Eqs. (4) and (5) in Randolph et al. (2008) for complete algorithms.

Algorithm	Equation	Citation	Type, location	Original sensor	# Ground-truthed data	Ground-truthed waterbody	Method
1 NDCI	$(b_4 - b_3)/(b_4 + b_3)$	Mishra and Mishra (2012)	semi-empirical, oceans	MERIS	56	Chesapeake Bay, Delaware Bay, the river Mississippi Delta region, and the Mobile Bay, USA	Calibrated with simulated data, validated with field data
2 SABI	$(b_4 - b_3)/(b_1 + b_2)$	Alawadi (2010)	semi-empirical, surface blooms	MODIS	0	–	Used Chlor- <i>a</i> MODIS product to validate
3 FLH blue	$(b_2) - [b_3 + (b_1 - b_3)]$	Zhao et al. (2010)	semi-empirical, oceans	Hyperspectral spectroradiometer	41	–	Validated with laboratory cultures
4 3BDA	$(b_3 - b_4) * b_4$	Dall'Omo and Gitelson (2005)	semi-empirical, reservoirs	Hyperspectral radiometer	144	2 sand pit lakes and 2 reservoirs, Nebraska; 1 lake, Iowa, USA	Calibrated with $N = 86$ and validation $N = 58$
5 KIVU (3BDA – like)	$(b_1 - b_3)/(b_2)$	Brivio et al. (2001)	semi-empirical, lakes	Landsat 5 TM	6	Lake Garda, Italy	Validated with in-situ data
6 Vincent	$(b_3/b_1) + (b_4/b_1) + (b_4/b_3) + (b_5/b_3) + (b_7/b_3) + (b_7/b_4)$	Vincent et al. (2004)	empirical, great lakes	Landsat 7 ETM+ and Landsat 5 TM	52	Lake Erie, USA	Calibrated model with in-situ data
7 Randolph (chl _a)	$((((R(709)/R(620))^*(0.727 + b_b) - b_b - 0.401) * (1/0.68)))$	Randolph et al. (2008)	semi-empirical, reservoirs	Hyperspectral spectroradiometer	55	Geist and Morse reservoirs, Indiana, USA	Calibrated model with in-situ data
8 Randolph (pcy)	$((((R(709)/R(620))^*(0.727 + b_b) - b_b - 0.281) * (1/0.84)) - (0.24 * a_{chl}))$	Randolph et al. (2008)	semi-empirical, reservoirs	Hyperspectral spectroradiometer	55	Geist and Morse reservoirs, Indiana, USA	Calibrated model with in-situ data
9 Cyanobacteria Index (CI)	$SS(\lambda) = R(\lambda) - R(\lambda^+) - [R(\lambda^+) - R(\lambda^-)] * [(R(\lambda) - R(\lambda^+))/(R(\lambda^+) - R(\lambda^-))]$ $(CI) = -SS(\lambda)$	Wynne et al. (2008)	empirical, lakes	Hyperspectral spectroradiometer, MERIS	NA	Bear Lake, MI, USA; Saginaw Bay, Lake Huron, MI, USA	Calibrated model with in-situ spectra, validated with satellite imagery
10 Cyanobacteria Index-multi (CI-multi)	If $-SS(665) < 0$, CI-multi = 0 If $-SS(665) > 0$, CI-multi = CI	Matthews et al. (2012)	empirical, oceans & lakes/reservoirs	MERIS	74	Benguela (Atlantic Ocean), Loskop Dam Reservoir (South Africa), Zeekoovlei (South Africa), Hartbeespoort Dam Reservoir (South Africa)	Validated with in-situ data

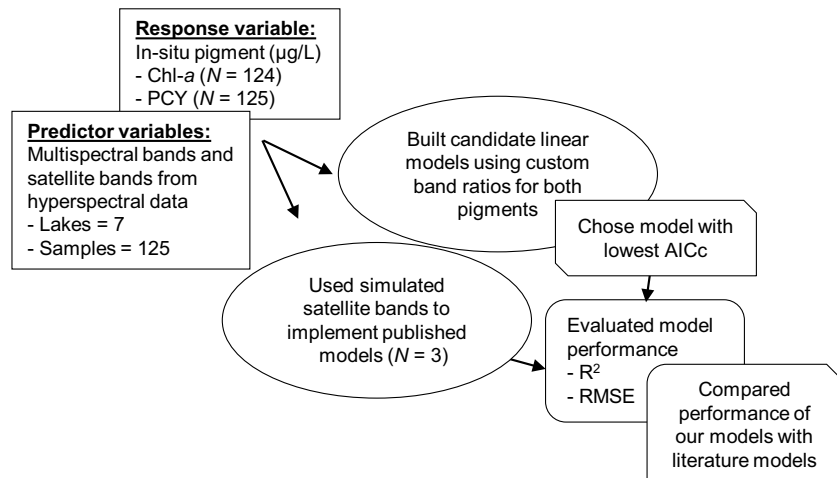
**Fig. 3.** Diagram showing ground-based model development methods. See methods for details.

Table 2

List of bands used to build a semi-empirical model to estimate phycocyanin. Bands were developed from hyperspectral data collected from Oklahoma lakes.

Band	Purpose	Minimum reflectance (nm)	Maximum reflectance (nm)
1	chlorophyll absorption minimum	546	591
2	phycocyanin trough	610	620
3	chlorophyll absorption trough	660	670
4	chlorophyll trough	678	684
5	reference band	700	710

including all possible combinations of ratios of the five bands and selected the best model using the corrected Akaike information criterion (AICc). As prediction was our goal, and not inferring causality, we moved forward in analyzing the model with the lowest AICc value. This process was repeated for phycocyanin. We compared our models (using adjusted R^2 and RMSE) with the chlorophyll-*a* and phycocyanin nested-band models from [Randolph et al. \(2008\)](#) which have been found to accurately predict chlorophyll-*a* and phycocyanin concentrations using hyperspectral data (see [Table 1](#) for algorithms and see Eqs. (4) and (5) in [Randolph et al. \(2008\)](#) for complete algorithms). We also compared our models with the Cyanobacterial Index (CI) and updated Cyanobacterial Index (CI-multi; [Matthews et al., 2012](#)) as defined in [Coffer et al. \(2020\)](#), which allows chlorophyll-*a* reflectance to be attributed to cyanobacteria based off the spectral shape at wavelength 665. To implement the CI and CI-multi algorithms, we first resampled our ground-based hyperspectral data to match the bands of MERIS by calculating the average reflectance over the wavelengths included in each band. Then we calculated CI and CI-multi following the methods of [Coffer et al. \(2020\)](#). The CI-multi algorithm is intended to filter out samples where phytoplankton composition is dominated by non-cyanobacteria. As such, we removed samples from further analysis (including model fit calculations) if the algorithm indicated cyanobacteria were not present. CI and CI-multi were converted to chlorophyll-*a* ($\mu\text{g/L}$) using the equation in ([Seegers et al., 2021](#)).

3. Results

3.1. Satellite models for chlorophyll-*a* and phycocyanin concentrations in small lakes

The best chlorophyll-*a* model built using Oklahoma data ($N = 1060$) contained four Landsat ETM+ and OLI band ratios: Blue:Green, Red:Blue, Blue:NIR, and NIR:Red ([Table 3.7](#)).

$$\text{Chl} - a = 71.8 - 37.3 (\text{Blue:Green}) - 10.1 (\text{Red:Blue}) - 6.6 (\text{Blue:NIR}) - 3.8 (\text{NIR:Red}), \quad (1)$$

Based on the adjusted R^2 , RMSE, and Pearson's r correlation coefficient, our Oklahoma-derived model outperformed all other models ([Table 3.1–6](#)), although they had relatively low predictive power. Of the literature algorithms, the KIVU algorithm (adapted from [Beck et al., 2016](#)) performed the best, having the lowest RMSE and highest adjusted R^2 values, although it had low predictability. The remainder of the literature models performed poorly with little to no predictive power for algal pigments in Oklahoma lakes ([Table 3.1–6](#)).

The initial performance of our chlorophyll-*a* model was poor. This poor performance could be due in part to the skewed nature of our data, with the majority of the chlorophyll-*a* values falling between 2 and 30 $\mu\text{g/L}$ ([Fig. S1](#)). To test if this skew in the data was causing poor prediction in the model we resampled chlorophyll-*a* data to create a more even distribution. Specifically, the original data was sorted into 30 intervals. The intervals containing the chlorophyll-*a* values 2–37 $\mu\text{g/L}$

(intervals 2–8) were randomly subsampled without replacement to reduce the whole data set by approximately half ($N = 511$) ([Fig. S2](#)). We retained all higher chlorophyll-*a* values, because while we believe the original data distribution is reflective of total chlorophyll-*a* values seen annually in Oklahoma lakes, we are interested in predicting the higher chlorophyll-*a* values that would be associated with a bloom. We then applied the best chlorophyll-*a* model ([Eq. \(1\)](#)) to the new data set ($N = 511$) and found the predictability did not improve from the full data set. The literature algorithms were also tested on the reduced dataset and there was no improvement in performance ([Table S4.1–6](#)). All models, including our own, showed systematic bias where samples with medium to high (30–150 $\mu\text{g/L}$) observed chlorophyll-*a* values had much lower predicted values, never exceeding 30 $\mu\text{g/L}$ ([Fig. 4A](#)).

The best phycocyanin ($N = 97$) model built using Oklahoma data used two band ratios: Blue:Green and Red:NIR ([Table 3.10](#)).

$$\text{PCY} = 146.7 - 124.5 (\text{Blue:Green}) - 16.7 (\text{Red:NIR}), \quad (2)$$

This phycocyanin model underpredicted at higher values, meaning it predicted much lower phycocyanin values than were observed ([Fig. 4B](#)). We also tested a phycocyanin model from the literature, the [Vincent et al. \(2004\)](#) model, on the Oklahoma dataset. The [Vincent et al. \(2004\)](#) model performed better than the model from this paper ([Table 3.9](#)).

The chlorophyll-*a* random forest model constructed using satellite data performed poorly, only explaining 3.07% of the variance ([Table 3.11](#)) and it was not significantly different when tested against a null model built with the same data. The phycocyanin random forest explained 32.9% of the variance and performed better than our linear two-band ratio model but not as well as Vincent's satellite phycocyanin model ([Table 3.9–10, 11](#)).

3.2. Ground-based models for chlorophyll-*a* and phycocyanin concentrations in small lakes

To test the efficacy of ground-based sensors at quantifying cyanoHABs in Oklahoma we built a semi-empirical multispectral model using hyperspectral data taken from Oklahoma lakes over a variety of conditions to predict chlorophyll-*a* and phycocyanin concentrations, respectively. The best chlorophyll-*a* model included six band ratios (1:3, 1:5, 2:3, 3:4, 3:5, 4:5; see [Table 2](#) for band information, see [Eq. S1](#) for full equation). The CI, CI-multi, and [Randolph et al. \(2008\)](#) chlorophyll-*a* nested-band-ratio model were also applied to the Oklahoma lakes data and resulted in slightly lower performance compared to our six-band ratio multispectral model ([Table 4.1–4; Fig. 5](#)).

Our best phycocyanin model included seven band ratios (1:2, 1:5, 2:3, 2:4, 3:4, 3:5, 4:5; see [Table 2](#) for band information, see [Eq. S2](#) for full equation). Comparison of the predictions of our phycocyanin seven-band ratio multispectral model with those of a common nested-band-ratio model ([Randolph et al., 2008](#)), CI, and CI-multi revealed increased accuracy with the seven-band ratio approach ([Table 4.5–8; Fig. 6](#)).

4. Discussion

The frequency and magnitude of cyanoHABs are increasing globally, in pace with climate and land-use change, and increasing nutrient pollution ([Huisman et al., 2018](#)). Detecting and tracking blooms in a timely manner for risk management has proven to be difficult and costly ([Almuhitaram et al., 2021](#)). Remote sensing has the potential to alleviate the insufficiencies of traditional monitoring by providing fast and less expensive information while allowing the end user to monitor many lakes simultaneously. Ground-based remote sensing has increased in popularity in recent years and appears more adaptable for many different systems and scenarios, thus it could be better for monitoring cyanoHABs. Here, we assessed the possibility of using satellite- and ground-based remote sensing for quantifying cyanoHABs in Oklahoma reservoirs.

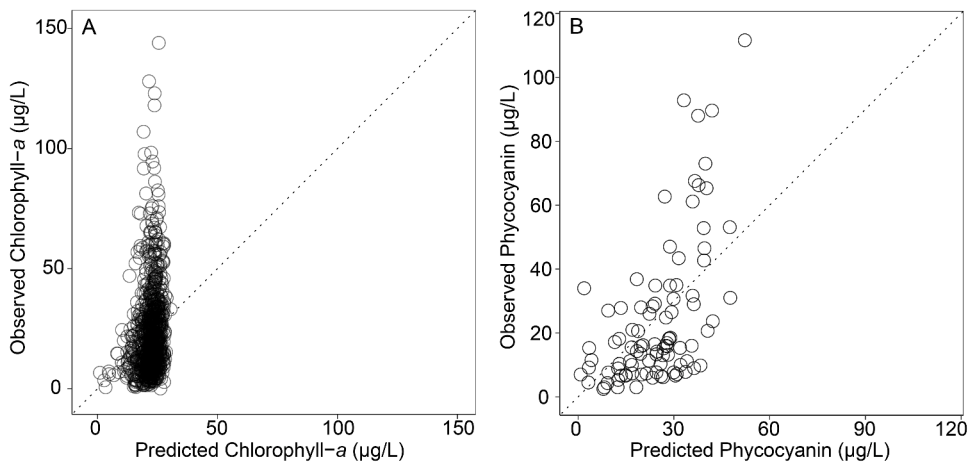


Fig. 4. Comparison of predicted and observed in-situ concentrations of algal pigments based on models built using Landsat 7 and 8 reflectance values. (A) The best chlorophyll-*a* ($N = 1060$) model based on BIC had an adjusted $R^2 = 0.035$ and $p \leq 0.001$. (B) The best phycocyanin ($N = 97$) model based on BIC had an adjusted $R^2 = 0.29$ and $p \leq 0.001$. Dashed line shows a 1:1 ratio between observed and predicted pigment concentrations. Darker areas in (A) indicate increased overlap of points.

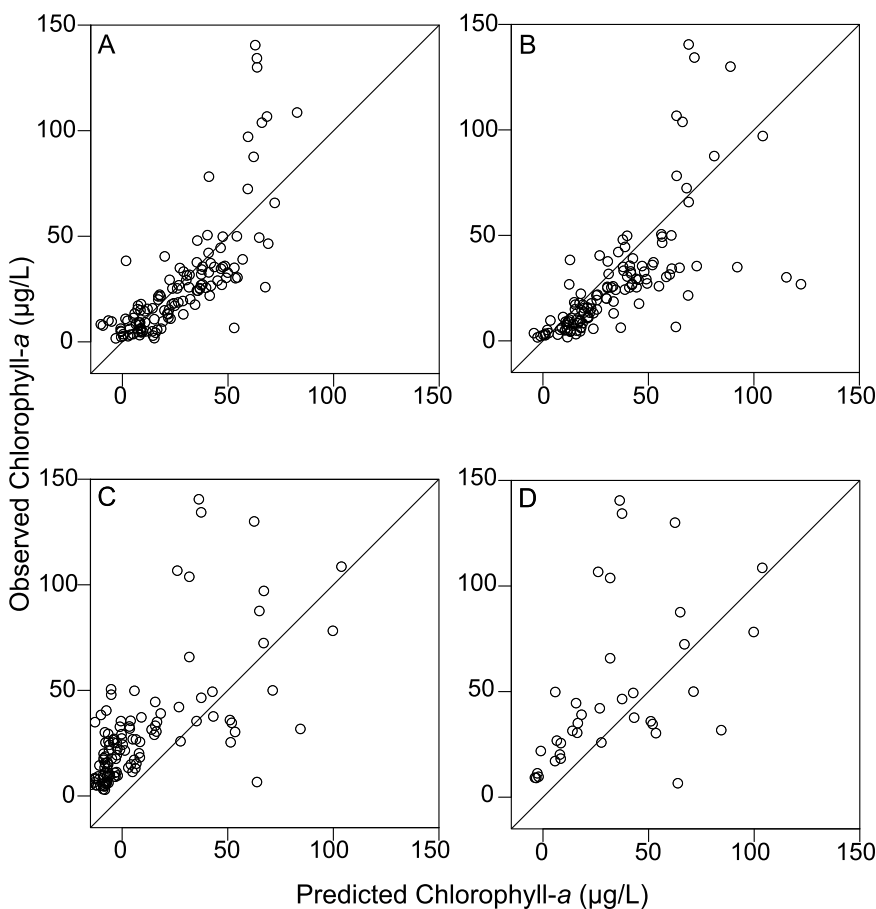


Fig. 5. Comparison of the observed in-situ chlorophyll-*a* and the predicted chlorophyll-*a* concentrations from (A) the Oklahoma multispectral 6-band ratio model from this paper ($N = 124$), (B) the Randolph model ($N = 124$), (C) the Cyanobacterial Index (CI) ($N = 124$), and (D) the updated Cyanobacterial Index (CI-multi) ($N = 37$). See Table 4 for model fit statistics. The solid line shows a 1:1 ratio between observed and predicted pigment concentrations to visually show the accuracy of the predicted values (closer to the line is more accurate). Darker areas indicate increased overlap of points.

We were unsuccessful in finding a ‘universal’ satellite pigment model for Oklahoma lakes. Our Landsat-based models for chlorophyll-*a* derived from Oklahoma reservoirs were only marginally better than published algorithms, and none of them were sufficiently predictive ($R^2 \leq 0.035$) for implementation in monitoring programs. Even though we were using mostly Landsat 7 data, our results corroborate those of Beck et al. (2016), who applied these algorithms to Landsat 8 simulated imagery and found poor performance across the board except with the Fluorescence Line Height violet algorithm. We were unable to use this algorithm because Landsat 7 lacks a comparable band to Landsat 8’s Coastal/Aerosol Band 1 (0.435–0.451 nm) used in the algorithm. While it was our ‘best’ performing literature algorithm, we expected better

performance from the KIVU 3-band-like algorithm because it performed well on Lake Garda when adapted from MERIS bands to Landsat 5 TM bands (Brivio et al., 2001), which are comparable to Landsat 7 ETM+ bands. The other algorithms were originally created using other sensors, such as MERIS and MODIS, or hyperspectral data that was translated to match Landsat satellite bands. This cross-sensor translation was likely another contributor to the poor performance of the models (Beck et al., 2016).

The poor prediction of the algorithm may also reflect the size of our data set, consisting of 1060 paired data points (much larger than previous studies) and a large number of waterbodies used ($N = 38$). The reflectance values we measured ranged from 0.0025 to 0.37 and Landsat

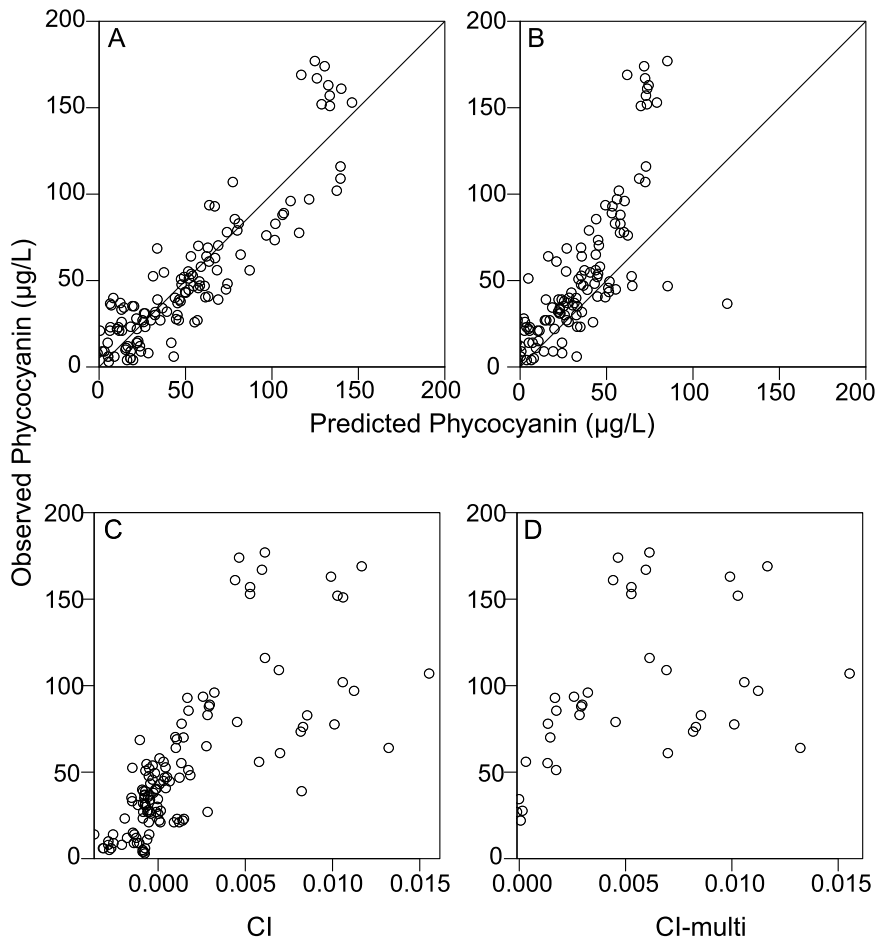


Fig. 6. Comparison of the observed and predicted concentrations of phycocyanin based on the Oklahoma multi-spectral 7-band ratio model ($N = 125$) from this paper (A), the Randolph model ($N = 125$) (B), CI (C), and CI-multi (D). Model fit statistics can be found in Table 4. The CI ($N = 125$) and CI-multi ($N = 38$) were used to 'predict' phycocyanin. There are currently no models translating the raw model output (CI, CI-multi) to $\mu\text{g/L}$ phycocyanin, therefore observed phycocyanin was plotted against the raw model outputs (unitless) to test correlation. The solid line in panels A and B shows a 1:1 ratio between observed and predicted pigment concentrations to visually show the accuracy of the predicted values (closer to the line is more accurate). Darker areas indicate increased overlap of points.

surface reflectance values ranged from 0 to 1, meaning instrument saturation is unlikely the cause of the poor performance (*Landsat surface reflectance data*, 2015). Instead, this prediction bias is likely due to the distribution of the original data, where the majority of the chlorophyll-*a* values ranged 2–30 $\mu\text{g/L}$. Considering the wide temporal nature of our data (2001–2017; all months retained), we believe this chlorophyll-*a* distribution is representative of natural conditions. Even when the data were subsampled to reduce the potential bias of abundant values in the low-medium chlorophyll-*a* range, the models did not improve in accurately predicting chlorophyll-*a*. The previously published models also displayed this bias, with no model accurately predicting values of chlorophyll-*a* greater than $\sim 30 \mu\text{g/L}$, and following a similar shape as displayed in Fig. 4A. With the lack of correspondence between high observed and predicted chlorophyll-*a* values it seems unlikely a universal Landsat algorithm will accurately predict chlorophyll-*a* concentrations in Oklahoma.

The phycocyanin satellite models performed better than the chlorophyll-*a* models and specifically, the phycocyanin model from Vincent et al. (2004) outperformed our model (Table 3). The better performance by the phycocyanin models was a surprising finding considering current Landsat sensor bands are not optimized for detecting phycocyanin. This pigment, found only in cyanobacteria, is characterized by an absorption trough around 621 nm (Almuh-taram et al., 2021). Due to the lack of a specific band for phycocyanin, cyanobacteria are detected using bands associated with chlorophyll-*a* when using Landsat satellites, making cyanobacteria frequently indistinguishable from aquatic macrophytes (Oyama et al., 2015). The detection of cyanobacteria is also confounded by turbidity, as bands commonly used in chlorophyll-*a* retrieval algorithms are susceptible to interference by suspended sediment (Almuh-taram et al., 2021; Shi et al., 2019). In fact, others have hypothesized the

Table 3

Performance of satellite algorithms for chlorophyll-*a* and phycocyanin estimation on Oklahoma lakes was evaluated using the adjusted R^2 from the given linear model, the root mean square error (RMSE), and Pearson's r correlation coefficient. *The RMSE for the resampled model should not be directly compared to the other models because the dataset used to build the model was different.

	Algorithm name	Adjusted R^2	Model P-value	RMSE	Pearson's r
Chlorophyll- <i>a</i> algorithms					
1	NDCI	−0.0009	0.81	17.57	0.007
2	SABI	0.001	0.16	17.55	0.044
3	FLHB	−0.0009	0.90	17.57	0.004
4	2BDA	0.002	0.10	17.55	0.050
5	3BDA	0.0003	0.25	17.56	0.035
6	KIVU (3BDA – like)	0.003	0.05	17.54	0.060
7	this paper (full dataset)	0.035	<0.001	17.22	0.198
8	this paper (resampled dataset)	0.025	<0.001	21.95*	0.170
Phycocyanin algorithms					
9	Vincent	0.384	<0.001	16.85	0.650
10	this paper	0.291	<0.001	18.49	0.553
	Random Forests	R²	% Variance explained	RMSE	Model significantly different from null?
11	chlorophyll- <i>a</i>	0.011	3.07	17.30	No
12	phycocyanin	0.321	32.86	18.18	No

Table 4

Performance of hyperspectral algorithms for chlorophyll-*a* and phycocyanin estimation on Oklahoma lakes evaluated using the adjusted R^2 from the given linear model, the root mean square error (RMSE), and Pearson's r correlation coefficient.

	Algorithm name	Adjusted R^2	Model P -value	RMSE	Pearson's r
Chlorophyll-<i>a</i> algorithms					
1	Randolph et al. model	0.509	<0.001	22.07	0.717
2	CI	0.472	<0.001	20.26	0.690
3	CI-multi	0.376	<0.001	21.93	0.617
4	this paper	0.660	<0.001	16.03	0.814
Phycocyanin algorithms					
5	Randolph et al. model	0.592	<0.001	34.99	0.772
6	CI	0.594	<0.001	69.33	0.773
7	CI-multi	0.519	<0.001	69.33	0.723
8	this paper	0.816	<0.001	19.16	0.904

Vincent et al. (2004) algorithm is simply detecting turbidity or chlorophyll-*a* which are correlated with phycocyanin (Hunter et al., 2010). The phycocyanin dataset ($N = 97$) was an order of magnitude smaller than chlorophyll-*a* dataset and only represented two lakes, Lake Thunderbird and Lake Texoma. The small number of lakes could contribute to the increased predictability in these models, potentially due to a decrease in noise by fewer optical characteristic combinations in the data, as discussed above. Since our phycocyanin model outperformed our chlorophyll-*a* model, we suggest implementing the phycocyanin model if satellites are the only possible source of spectral data. Unlike chlorophyll-*a*, phycocyanin has been shown to be well correlated with cyanobacterial biomass (Thomson-Lang 2020).

We utilized multiple commonly used metrics for determining model performance of the satellite models, including R^2 , RMSE, and Pearson's R . The next step would be to validate these models by collecting additional ground-reference data and using the independent dataset to measure model performance. We did not do this for the chlorophyll-*a* and phycocyanin Landsat models simply due to their poor performance on the initial data pool.

As for our random forest models built using satellite data, the chlorophyll-*a* random forest regression model performed better than all of the literature models but slightly worse than our satellite linear model. The phycocyanin random forest model was better than our model but not better than the best phycocyanin model, the Vincent model. We thought because random forests are based on non-linear models that they would outperform the multiple linear regression models in our dataset, but this was not the case. While random forest modeling is a newer, machine learning approach that might be touted as a better solution to model problems, one of the drawbacks is the difficulty of implementation of the model, as you cannot 'see' the model or report a formula describing the forest. Trying more 'sophisticated' modeling techniques may not be the correct solution if your underlying data is not showing any predictive trends using simpler modeling methods.

Many of the current standard algorithms were calibrated or validated on very few ground-reference data points from a small number of lakes. Literature algorithms were calibrated on five or fewer lakes. The number of unique optical characteristics, such as different colors and levels of turbidity and suspended solids, would likely increase with lake number, thus increasing the overall variability in satellite reflectance values. One of the obvious goals of universal algorithms is to predict cyanoHABs across many different lakes and seasons, therefore we think it is important to include a wide range of training data for model building. The lack of representative training data could be why the published satellite-based algorithms, for which training data does not include data from Oklahoma lakes, perform worse than our model on Oklahoma lakes. Given the failure of satellite-based models in predicting chlorophyll-*a* within a single U.S. state, we are doubtful that a truly universal

model can be developed.

The disconnect between satellite observations and actual pigment concentrations could also be due to satellite sensors not capturing all water-leaving reflectance due to atmospheric interference or cloud-cover, and not capturing data below the surface of the water column, because the satellite sensors only measure chlorophyll-*a* at the surface, but *in-situ* measurements of chlorophyll-*a* are generally taken across the photic zone for mid-column bloom formers. This would cause mid-column blooming cyanobacteria or well mixed blooms to appear less concentrated or less pigmented than surface blooms (Coffer et al., 2021a). In Oklahoma lakes, for example, we frequently experience *Raphidiopsis* blooms and other non-surface bloomers (Antunes et al., 2015). These problems would be less of an issue for ground-based sensors that are not affected by atmospheric aerosols and when suspended close to the water surface can capture a large majority of the water leaving reflectance.

Our ground-based multispectral models reliably predicted both chlorophyll-*a* and phycocyanin in Oklahoma lakes and performed remarkably better than the satellite models. The chlorophyll-*a* six-band ratio model we developed slightly outperformed Randolph et al. (2008) nested-band ratio chlorophyll-*a* model, with both models having similar R^2 values. The CI and CI-multi performed substantially worse than our model and Randolph et al. (2008) model. All four models generally underpredicted at very high levels of chlorophyll-*a*. Our phycocyanin seven-band ratio model outperformed the phycocyanin nested-band ratio model from Randolph et al. (2008), the CI model, and CI-multi model. Notably, our model performed better on extreme phycocyanin values ($> 50 \mu\text{g/L}$) compared with the established model which had systematic underprediction of high phycocyanin events. While Coffer et al. (2021b, 2021a, 2020) and Handler et al. (2023) have successfully used MERIS and OLCI data to detect cyanobacteria in larger Oklahoma reservoirs, we did not corroborate these results with our simulated data. The CI-multi includes a filtering step with the aim of removing non-cyanobacterial blooms after calculating the CI (Coffer et al., 2020). This filtering step is meant to remove samples that are not from cyanobacteria dominated systems, i.e., classified as CI-noncyano. This step seemed to fail in our system, with 87 samples filtered out. With the exception of one high outlier, the phycocyanin in these samples ranged between 3 and $71 \mu\text{g/L}$ phycocyanin, which are low to moderate levels of phycocyanin. We know that many of the lakes removed at this step are cyanobacteria dominated systems. For example, both Lake Texoma and Grand Lake O' the Cherokees have a rich history of cyanobacteria blooms. This failure of the CI-multi algorithm is problematic because predicting these low to moderate phycocyanin concentrations allows us to detect blooms as they develop. If detected early, measures may be taken to lower risks to humans, like closing swim beaches and posting educational signs. False negatives (model returning no bloom when blooms are present) could lead to failure to act to protect the public, pets, and livestock.

Our ground-based models, including the simulated CI models, outperformed all of the Landsat models. The poor performance of Landsat models is possibly due in part to the Landsat's band limitations. With limited band number and wavelength range, we found that even with a variety of algorithms employing different band math strategies, Landsat was not useful for detecting pigments in reservoirs. Researchers have moved away from using band ratios in recent years in favor of spectral shape algorithms, such as CI, or maximum peak height algorithms (Coffer et al., 2020; Matthews and Odermatt, 2015). We did find that our satellite models relying on band ratios were substantially worse than the CI models. Our results suggest that satellites where the CI could be implemented, such as Sentinel-3, would be better able to detect cyanobacteria blooms than Landsat. In Oklahoma, we found that ground-based instruments show more promise than satellite-based instruments for monitoring cyanoHABs, but other methods exist for cyanobacteria bloom detection with different accuracy, scale, and cost trade-offs. These alternative approaches include various satellite

sensors, ground-based remote sensing, airplane-based sensors, and fluorometry options. Satellites with freely available images excel in the category of low cost, but as we have demonstrated, they lack accuracy and do not sample at spatial scales relevant to small or dendritic reservoirs (Fig. S6). Additionally, clouds are often an underestimated problem in collecting regular satellite images, which may limit timely detection of algal blooms (Ju and Roy 2008). As we have shown, ground-based remote sensing offers high accuracy at a fine scale but has an upfront cost for multi- or hyper-spectral sensors. As a bonus, such sensors could be deployed aerially, giving broader coverage. For example, the National Ecological Observatory Network (NEON) has deployed airplane-based hyperspectral sensors at NEON sites to measure land-cover and vegetation metrics at a fine scale without the interference of clouds (<https://www.neonscience.org/data-collection/imaging-spectrometer>). One trade-off of ground-based sensors is they are spatially limited compared to satellites. A multispectral sensor suspended above a lake would cover a very small area. A sensor attached to a drone or airplane would afford increased spatial coverage at the limitation of drone battery life or airplane fly time. Another option, deployed sondes, have high up-front and maintenance costs but provide highly accurate measurements on a fine scale. Manual sampling of water followed by in-lab extraction of chlorophyll has a high cost of human-hours but is accurate and operates on a fine scale (Almuharam et al., 2021). Given our results, we suggest that ground-based remote sensing offers the best of both worlds in terms of cost, accuracy, and scale for many scenarios, however the needs of the researcher or water quality manager will determine which of these approaches best meet their needs based on cost, accuracy, and scale.

Our finding that models using ground-based multispectral data better predict chlorophyll-*a* and phycocyanin concentrations compared to models built using Landsat imagery is likely applicable to small lakes and reservoirs around the globe, though transferring models such as ours to other systems may require additional ground-reference data and model adjustment. That is, we do not propose our model as a 'universal' model, but rather we suggest our approach of ground-based remote sensing coupled with custom spectral bands and band ratios. The lakes we analyzed in Oklahoma range from 49 to 42,695 hectares in size (Table S1) and are representative of smaller lakes and reservoirs across the southern Great Plains. There are millions of small lakes in the world (Verpoorter et al., 2014; Cael and Seekell, 2016) and reservoirs cover approximately 0.26 Mkm² globally (Downing et al., 2006). Many of these small waterbodies experience cyanoHABs that cannot be clearly resolved by larger satellite pixels. Because of this resolution limitation, small lakes and reservoirs are likely underrepresented in remote sensing monitoring programs. While Landsat has the appropriate pixel size for sensing small waterbodies, it does not have a band useful for detecting phycocyanin. As previously concluded by Beck et al. (2016), future satellite-based remote sensing for cyanoHABs will require higher resolution (30-m pixels or less), and more appropriate bands that are narrower and similar to those of WorldView-2/-3 and Sentinel-3. In the meantime, ground-based sensors offer an excellent alternative for small- to mid-sized lakes. Timely and effective monitoring of water quality in small to moderate size lakes cannot rely on satellite sensors with inappropriate spatial resolution and spectral band design. Instead, we recommend the novel and herein demonstrated successful approach of ground-based remote sensing with customized bands. Our approach will be a valuable addition to water quality monitoring efforts in small and dendritic reservoirs.

5. Conclusion

Based on a multi-year, multi-lake comparison of *in-situ* algal pigment data with Landsat- and ground-derived reflectance models, we conclude that:

- Oklahoma-derived Landsat- and ground-based models outperform established reflectance-pigment models for Oklahoma reservoirs.
- Ground-based, multispectral models are superior to Landsat-based models for predicting cyanoHABs in Oklahoma reservoirs.
- Ground-based and multispectral sensors can offer cost-efficient solutions for cyanoHAB monitoring in small- to mid-sized lakes where satellite images may not be appropriate.

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.

Data availability

Data will be made available on request.

Acknowledgments

Landsat Surface Reflectance products are courtesy of the U.S. Geological Survey Earth Resources Observation and Science Center. Special thanks to: Karen Glenn, James Easton, Anne Easton and Thayer Hallidayshult for sample collection and processing; Brandon Chien, Dani Glidewell, Lena Wilson, Mary Larson, and Garrett Eakers for image collection and processing; the University of Oklahoma Biological Station staff for facilities support; Jie Wang and Jinheng Zhang for sample collection and data processing, and Rajen Bajgain, Yuting Zhou, and Zhenhua Zou for technical assistance; Julie Chambers and Chris Adams for providing pigment data; Rich Zamor and Steve Nikolai for sampling assistance and providing data; Andy Dzialowski and Keith Loftin for early discussions. This work will be submitted by KVC as partial fulfillment of the requirements for a Ph.D. degree from the University of Oklahoma; additional thanks to her committee members, Drs. Alan Wilson, Katharine Marske, and Amy McGovern, for their support in the preparation of this manuscript. This research was supported in part by grants from the Oklahoma Water Resources Research Institute and the Oklahoma Water Resources Board (Agreement No. 2013OK292B-OU1/2), the Oklahoma Department of Wildlife Conservation (through the Sport Fish Restoration Program, Grant F-61-R), NSF (DEB 1831061, OIA-1946093), USGS (G21AP10181), and an NSF Graduate Research Fellowship.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.watres.2023.120076](https://doi.org/10.1016/j.watres.2023.120076).

References

- Alawadi, F., 2010. Detection of surface algal blooms using the newly developed algorithm surface algal bloom index (SABI). *Int. Soc. Opt. Photonics*, 782506. <https://doi.org/10.1117/12.862096>.
- Almuharam, H., Kibuye, F.A., Ajampur, S., Glover, C.M., Hofmann, R., Gaget, V., Owen, C., Wert, E.C., Zamyadi, A., 2021. State of knowledge on early warning tools for cyanobacteria detection. *Ecol. Indic.* 133, 108442 <https://doi.org/10.1016/j.ecolind.2021.108442>.
- Antunes, J.T., Leão, P.N., Vasconcelos, V.M., 2015. *Cylindrospermopsis raciborskii*: review of the distribution, phylogeography, and ecophysiology of a global invasive species. *Front. Microbiol.* 6, 473. <https://doi.org/10.3389/fmicb.2015.00473>.
- Arar, E.J., Collins, G.B., 1997. Method 445.0 In Vitro Determination of Chlorophyll *a* and Pheophytin *a* in Marine and Freshwater Algae by Fluorescence. U.S. Environmental Protection Agency, Washington, DC.
- Baban, S.M.J., 1993. Detecting water quality parameters in the Norfolk Broads, U.K., using Landsat imagery. *Int. J. Remote Sens.* 14, 1247–1267. <https://doi.org/10.1080/01431169308953955>.
- Barnes, W.L., Pagano, T.S., Salomonson, V.V., 1998. Prelaunch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS-AM1. *IEEE Trans. Geosci. Remote Sens.* 36, 1088–1100. <https://doi.org/10.1109/36.700993>.

- Beck, R., Zhan, S., Liu, H., Tong, S., Yang, B., Xu, M., Ye, Z., Huang, Y., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Nietch, C., Macke, D., Martin, M., Stillings, G., Stump, R., Su, H., 2016. Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. *Remote Sens. Environ.* 178, 15–30. <https://doi.org/10.1016/j.rse.2016.03.002>.
- Beyer, J.E., Hambright, K.D., 2017. Maternal effects are no match for stressful conditions: a test of the maternal match hypothesis in a common zooplankton. *Funct. Ecol.* 31, 1933–1940. <https://doi.org/10.1111/1365-2435.12901>.
- Brivio, P.A., Giardino, C., Zilioli, E., 2001. Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images. *Int. J. Remote Sens.* 22, 487–502. <https://doi.org/10.1080/014311601450059>.
- Cael, B.B., Seekell, D.A., 2016. The size-distribution of Earth's lakes. *Sci. Rep.* 6, 29633. <https://doi.org/10.1038/srep29633>.
- Coad, P., Cathers, B., Ball, J.E., Kadluczka, R., 2014. Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial neural network. *Environ. Model. Softw.* 61, 393–409. <https://doi.org/10.1016/j.envsoft.2014.07.011>.
- 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>.
- Coffer, M.M., Schaeffer, B.A., Foreman, K., Porteous, A., Loftin, K.A., Stumpf, R.P., Werdell, P.J., Urquhart, E., Albert, R.J., Darling, J.A., 2021a. Assessing cyanobacterial frequency and abundance at surface waters near drinking water intakes across the United States. *Water Res.* 201, 117377. <https://doi.org/10.1016/j.watres.2021.117377>.
- Coffer, M.M., Schaeffer, B.A., Salls, W.B., Urquhart, E., Loftin, K.A., Stumpf, R.P., Werdell, P.J., Darling, J.A., 2021b. Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales. *Ecol. Indic.* 128, 107822. <https://doi.org/10.1016/j.ecolind.2021.107822>.
- Dall'Olmo, G., Gitelson, A.A., 2005. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Appl. Opt.*, AO 44, 412–422. <https://doi.org/10.1364/AO.44.000412>.
- Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M., Middelburg, J.J., 2006. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* 51, 2388–2397. <https://doi.org/10.4319/lo.2006.51.5.2388>.
- Fernandez-Figueroa, E.G., Wilson, A.E., Rogers, S.R., 2021. Commercially available unoccupied aerial systems for monitoring harmful algal blooms: a comparative study. *Limnol. Oceanogr. Methods*. <https://doi.org/10.1002/lom3.10477>.
- Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* 160, 271–282. <https://doi.org/10.1078/0176-1617-00887>.
- Gons, H.J., Rijkeboer, M., Ruddick, K.G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *J. Plankton Res.* 24, 947–951. <https://doi.org/10.1093/plankt/24.9.947>.
- Hambright, K.D., Zamor, R.M., Easton, J.D., Allison, B., 2014. Algae. *Encyclopedia of Toxicology*. Elsevier, pp. 130–141. <https://doi.org/10.1016/B978-0-12-386454-3.00983-0>.
- Hambright, K.D., Zamor, R.M., Easton, J.D., Glenn, K.L., Rummel, E.J., Easton, A.C., 2010. Temporal and spatial variability of an invasive toxigenic protist in a North American subtropical reservoir. *Harmful Algae* 9, 568–577. <https://doi.org/10.1016/j.hal.2010.04.006>.
- Handler, A.M., Compton, J.E., Hill, R.A., Leibowitz, S.G., Schaeffer, B.A., 2023. Identifying lakes at risk of toxic cyanobacterial blooms using satellite imagery and field surveys across the United States. *Sci. Total Environ.* 869, 161784. <https://doi.org/10.1016/j.scitotenv.2023.161784>.
- Hilborn, E.D., Beasley, V.R., 2015. One health and cyanobacteria in freshwater systems: animal illnesses and deaths are sentinel events for human health risks. *Toxins (Basel)* 7, 1374–1395. <https://doi.org/10.3390/toxins7041374>.
- Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* 16, 471–483. <https://doi.org/10.1038/s41579-018-0040-1>.
- Hunter, P.D., Tyler, A.N., Carvalho, L., Codd, G.A., Maberly, S.C., 2010. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. *Remote Sens. Environ.* 114, 2705–2718. <https://doi.org/10.1016/j.rse.2010.06.006>.
- Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sens. Environ.*, Landsat Legacy Special Issue 122, 11–21. <https://doi.org/10.1016/j.rse.2011.08.026>.
- Ju, J., Roy, D.P., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens. Environ.* 112, 1196–1211. <https://doi.org/10.1016/j.rse.2007.08.011>.
- Kneubühler, M., Frank, T., Kellenberger, T., Pasche, N., Schmid, M., 2007. Mapping chlorophyll-a in Lake Kivu with remote sensing methods, in: Lacoste, H., Ouwehand, L. (Eds.). European Space Agency * Communication Production Office, Montreux (CH).
- Landsat surface reflectance data (Report No. 2015–3034), 2015, Fact Sheet. Reston, VA. <https://doi.org/10.3133/fs20153034>.
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2, 18–22.
- Markham, B.L., Storey, J.C., Williams, D.L., Irons, J.R., 2004. Landsat sensor performance: history and current status. *IEEE Trans. Geosci. Remote Sens.* 42, 2691–2694. <https://doi.org/10.1109/TGRS.2004.840720>.
- Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* 3, 68–72. <https://doi.org/10.1109/LGRS.2005.857030>.
- Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* 32, 6855–6899. <https://doi.org/10.1080/01431161.2010.512947>.
- Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* 124, 637–652. <https://doi.org/10.1016/j.rse.2012.05.032>.
- Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. *Remote Sens. Environ.* 156, 374–382. <https://doi.org/10.1016/j.rse.2014.10.010>.
- Mishra, S., Mishra, D.R., 2012. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sens. Environ./Remote Sens. Urban Environ.* 117, 394–406. <https://doi.org/10.1016/j.rse.2011.10.016>.
- Nieke, J., Borde, F., Mavrocordatos, C., Berruti, B., Delclaud, Y., Riti, J.B., Garnier, T., 2012. The ocean and land colour imager (OLCI) for the sentinel 3 GMES mission: status and first test results. In: *International Society for Optics and Photonics*, p. 85280C.
- Oklahoma Water Resources Board, 2017. Beneficial Use Monitoring Program (Lake Report). Oklahoma Water Resources Board, State of Oklahoma, Oklahoma City, OK.
- Overstreet, B.T., Legleiter, C.J., 2017. Removing sun glint from optical remote sensing images of shallow rivers. *Earth Surf. Process. Landforms* 42, 318–333. <https://doi.org/10.1002/esp.4063>.
- Oyama, Y., Matsushita, B., Fukushima, T., 2015. Distinguishing surface cyanobacterial blooms and aquatic macrophytes using Landsat/TM and ETM+ shortwave infrared bands. *Remote Sens. Environ.*, Special Issue: Remote Sens. Inland Waters 157, 35–47. <https://doi.org/10.1016/j.rse.2014.04.031>.
- Paerl, H.W., Barnard, M.A., 2020. Mitigating the global expansion of harmful cyanobacterial blooms: moving targets in a human- and climatically-altered world. *Harmful Algae* 96, 101845. <https://doi.org/10.1016/j.hal.2020.101845>.
- Paerl, H.W., Paul, V.J., 2012. Climate change: links to global expansion of harmful cyanobacteria. *Water Res.* 46, 1349–1363. <https://doi.org/10.1016/j.watres.2011.08.002>.
- Plaas, H.E., Paerl, H.W., 2021. Toxic cyanobacteria: a growing threat to water and air quality. *Environ. Sci. Technol.* 55, 44–64. <https://doi.org/10.1021/acs.est.0c06653>.
- R Development Core Team, 2014. R: A Language and Environment For Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Randolph, C., Wilson, J., Tedesco, L., Li, L., Pascual, D.L., Soyeux, E., 2008. Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin. *Remote Sens. Environ.*, Appl. Remote Sens. Monit. Freshwater Estuarine Syst. 112, 4009–4019. <https://doi.org/10.1016/j.rse.2008.06.002>.
- Rast, M., Bezy, J.L., Bruzzi, S., 1999. The ESA medium resolution imaging spectrometer MERIS a review of the instrument and its mission. *Int. J. Remote Sens.* 20, 1681–1702. <https://doi.org/10.1080/014311699212416>.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Statist.* 6, 461–464. <https://doi.org/10.1214/aos/1176344136>.
- Seegers, B.N., Werdell, P.J., Vandermeulen, R.A., Salls, W., Stumpf, R.P., Schaeffer, B.A., Owens, T.J., Bailey, S.W., Scott, J.P., Loftin, K.A., 2021. Satellites for long-term monitoring of inland U.S. lakes: the MERIS time series and application for chlorophyll-a. *Remote Sens. Environ.* 266, 112685. <https://doi.org/10.1016/j.rse.2021.112685>.
- Shi, K., Zhang, Y., Qin, B., Zhou, B., 2019. Remote sensing of cyanobacterial blooms in inland waters: present knowledge and future challenges. *Sci. Bull.* 64, 1540–1556. <https://doi.org/10.1016/j.scib.2019.07.002>.
- Smith, D., Cauthron, B., Wright, J., Armstrong, C., Gilliland, G., Clyde, T., McCrackin, C., Hambright, K.D., Lynch, R., Moore, D., West, R., Phillips, E., Townsend, D., Ziara, S., 2012. Protocols for Harmful Algal Bloom Monitoring in Oklahoma Lakes. Office of the Secretary of Environment, Oklahoma City.
- 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>.
- Thomson-Laing, G., Puddick, J., Wood, S.A., 2020. Predicting cyanobacterial biovolumes from phycocyanin fluorescence using a handheld fluorometer in the field. *Harmful Algae* 97, 101869. <https://doi.org/10.1016/j.hal.2020.101869>.
- 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>.
- Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* 41, 6396–6402. <https://doi.org/10.1002/2014GL060641>.
- Vincent, R.K., Qin, X., McKay, R.M.L., Miner, J., Czajkowski, K., Savino, J., Bridgeman, T., 2004. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sens. Environ.* 89, 381–392. <https://doi.org/10.1016/j.rse.2003.10.014>.
- Wetzel, R.G., 2001. *Limnology: Lake and River Ecosystems*. Gulf Professional Publishing.

- Wu, D., Li, R., Zhang, F., Liu, J., 2019. A review on drone-based harmful algae blooms monitoring. *Environ. Monit. Assess.* 191, 211. <https://doi.org/10.1007/s10661-019-7365-8>.
- Wynne, T.T., Stumpf, R.P., Tomlinson, M.C., 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. *International Journal of Remote Sensing* 29 3665–3672. <https://doi.org/10.1080/01431160802007640>.
- Zhao, D., Xing, X., Liu, Y., Yang, J., Wang, L., 2010. The relation of chlorophyll-a concentration with the reflectance peak near 700 nm in algae-dominated waters and sensitivity of fluorescence algorithms for detecting algal bloom. *Int. J. Remote Sens.* 31, 39–48. <https://doi.org/10.1080/01431160902882512>.
- Zou, Z., Dong, J., Menarguez, M.A., Xiao, X., Qin, Y., Doughty, R.B., Hooker, K.V., David Hambright, K., 2017. Continued decrease of open surface water body area in Oklahoma during 1984–2015. *Sci. Total Environ.* 595, 451–460. <https://doi.org/10.1016/j.scitotenv.2017.03.259>.
- Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* 1, 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>.

Supplemental Information

Cook, K.V, Beyer, J.E, Xiangming, X, and Hambright, K.D. Ground-based remote sensing provides alternative to satellites for monitoring cyanobacterial blooms in small lakes (2023).

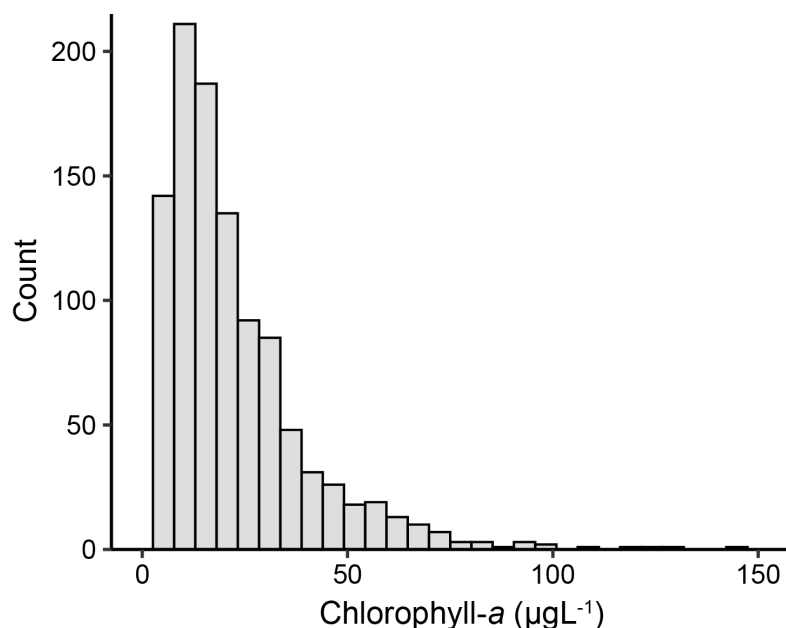


Figure S1. Distribution of chlorophyll-*a* in the full satellite dataset (n = 1160).

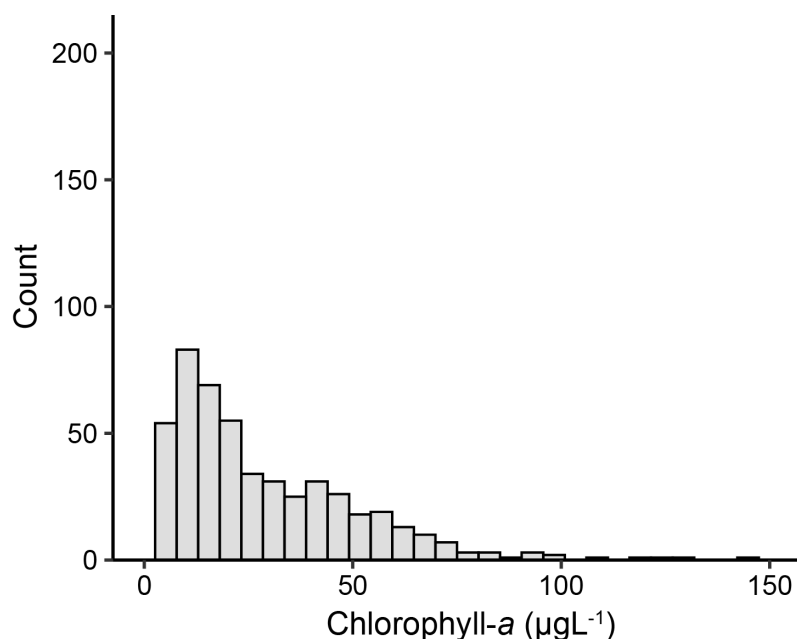


Figure S2. Distribution of chlorophyll-*a* in the re-sampled dataset (n = 511).

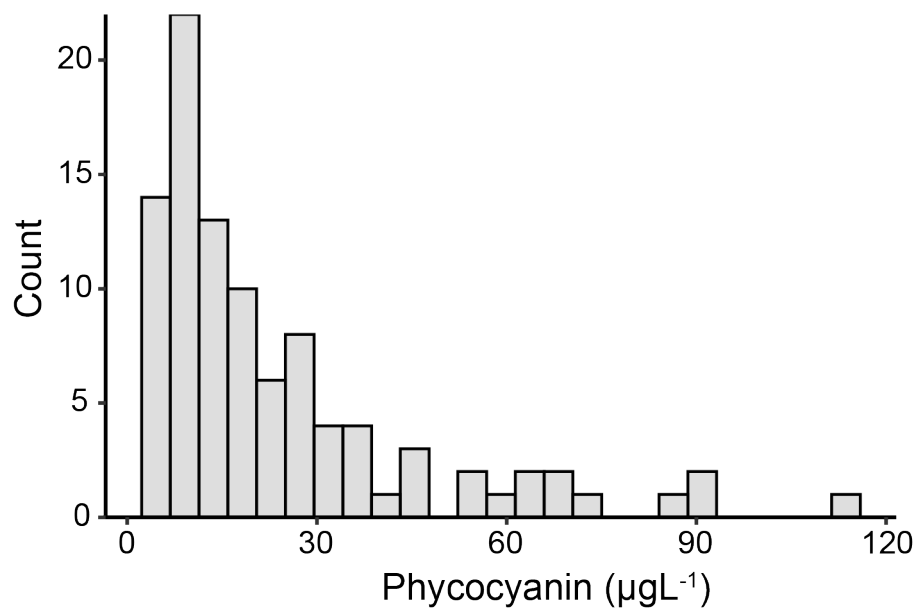


Figure S3. Distribution of phycocyanin in the full satellite dataset (n = 97).

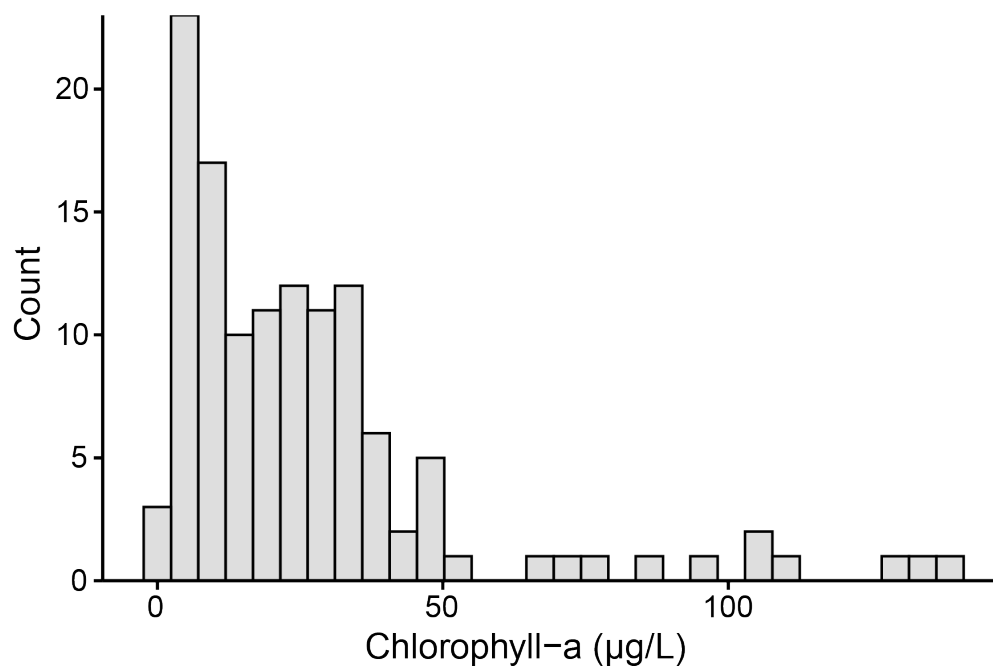


Figure S4. Distribution of chlorophyll-*a* in the ground-based dataset (n = 124).

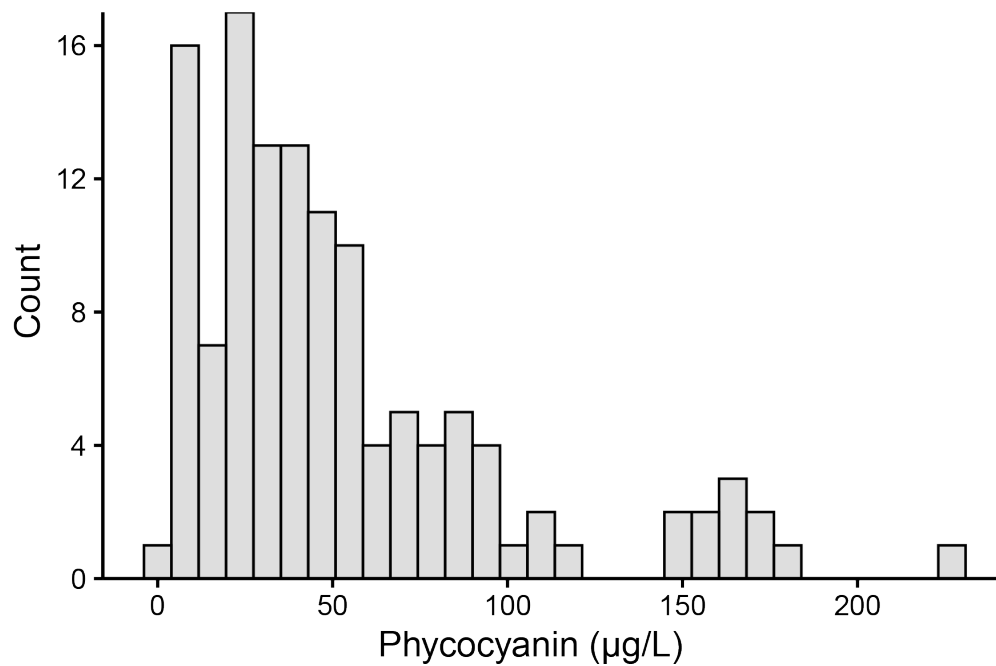


Figure S5. Distribution of phycocyanin in the ground-based dataset (n = 125).

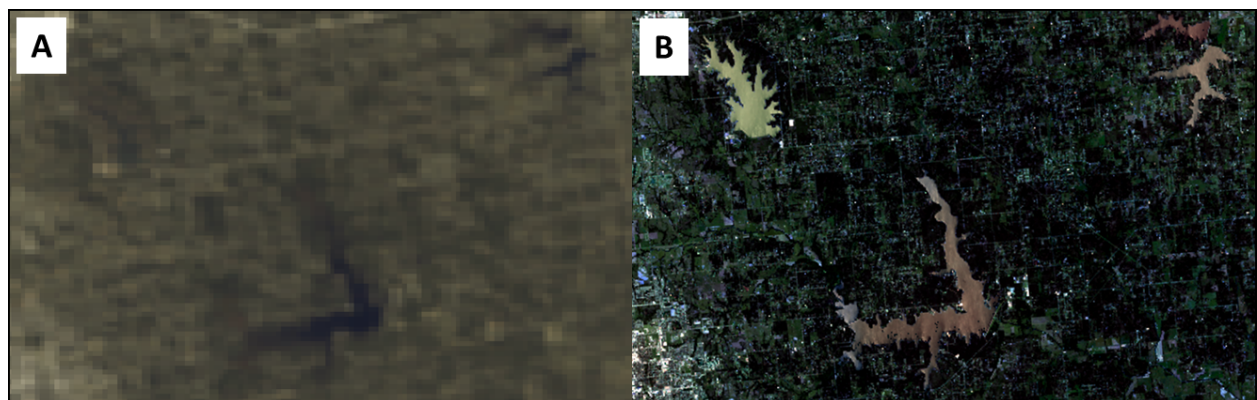


Figure S6. Comparison of Sentinel-3a and Landsat 8 images to demonstrate how the difference in pixel size impacts lake resolvability. Lake Thunderbird is pictured here at the center of each image, the Shawnee Twin Lakes in the upper right-hand corner, and Lake Stanley Draper to the upper left-hand corner of the images. (A) shows a Sentinel-3b (pixel size 300 x 300 m) false color composite of bands 1, 2, and 3 built using Sentinels Application Platform (SNAP) from the European Space Agency (ESA) and (B) shows a Landsat-8 (pixel size 30 x 30 m) false color composite of bands 2, 3, and 4 built in QGIS.

Table S1. Site information for the ground-truthed satellite chlorophyll-*a* data. Sampling organizations included Oklahoma Water Resources Board (OWRB), Plankton Ecology and Limnology Lab (PELL), and Grand River Dam Authority (GRDA).

Lake	Site	Number of samples	Sampling organization	Latitude	Longitude	Lake area (hectare)
Altus	Site 2	3	OWRB	34.90247	-99.2937	2,600
Altus	Site 3	2	OWRB	34.92733	-99.3130	
Altus	Surface	2	OWRB	34.88722	-99.2949	
Arcadia	Site 2	5	OWRB	35.63511	-97.3714	737
Arcadia	Site 3	5	OWRB	35.64489	-97.3853	
Arcadia	Site 4	5	OWRB	35.62770	-97.3939	
Arcadia	Site 5	5	OWRB	35.61594	-97.4059	
Atoka	Site 2	6	OWRB	34.47904	-96.0908	2,307
Atoka	Site 3	5	OWRB	34.50523	-96.0754	
Atoka	Site 4	3	OWRB	34.54017	-96.0462	
Atoka	Surface	3	OWRB	34.44624	-96.0872	
Birch	77	1	OWRB	36.51479	-96.1912	460
Canton	Site 2	2	OWRB	36.12572	-98.6014	3,201
Canton	Site 3	3	OWRB	36.14255	-98.6339	
Canton	Site 4	1	OWRB	36.14261	-98.6549	
Canton	Site 5	1	OWRB	36.09548	-98.5867	
Canton	Surface	3	OWRB	36.09337	-98.5883	
Chickasha	Site 2	4	OWRB	35.14737	-98.1502	837
Chickasha	Site 3	4	OWRB	35.14995	-98.1307	
Chickasha	Site 4	1	OWRB	35.13559	-98.1402	
Chickasha	Site 5	1	OWRB	35.14373	-98.1368	
Chickasha	Surface	4	OWRB	35.13199	-98.1329	
Claremore	Site 2	4	OWRB	36.34031	-95.5725	190
Claremore	Site 3	4	OWRB	36.34092	-95.5599	
Claremore	Site 4	2	OWRB	36.33351	-95.5786	
Claremore	Surface	4	OWRB	36.32514	-95.5796	
Clinton	Site 3	1	OWRB	35.42906	-99.2231	136
Copan	Site 2	4	OWRB	36.91109	-95.9564	1,963
Copan	Site 3	4	OWRB	36.92622	-95.9528	
Copan	Site 4	4	OWRB	36.94839	-95.9401	
Copan	Site 5	4	OWRB	36.93934	-95.9559	
Copan	Surface	4	OWRB	36.88641	-95.9679	
Crowder	Site 2	11	OWRB	35.39980	-98.7074	64
Cushing	80	2	NLA	36.00416	-96.879	170
Dave Boyer	73	2	NLA	34.37438	-98.3367	49

El Reno	2	5	PELL	35.52391	-97.9897	69
El Reno	3	5	PELL	35.51955	-97.9962	
Ellsworth	Site 2	5	OWRB	34.82185	-98.356	2,266
Ellsworth	Site 4	4	OWRB	34.81763	-98.3397	
Ellsworth	Site 5	3	OWRB	34.84256	-98.3584	
Eufaula	Site 11	6	OWRB	35.22886	-95.6338	42,695
Eufaula	Site 12	5	OWRB	35.20047	-95.5938	
Eufaula	Site 13	5	OWRB	35.16440	-95.5995	
Eufaula	Site 14	5	OWRB	35.10144	-95.6472	
Eufaula	Site 15	4	OWRB	35.05004	-95.6711	
Eufaula	Site 16	4	OWRB	35.01996	-95.6023	
Eufaula	Site 17	4	OWRB	34.97471	-95.6303	
Eufaula	Site 2	6	OWRB	35.42767	-95.6001	
Eufaula	Site 3	5	OWRB	35.38221	-95.6300	
Eufaula	Site 4	6	OWRB	35.30045	-95.5540	
Eufaula	Site 5	4	OWRB	35.28483	-95.5147	
Eufaula	Site 6	5	OWRB	35.30704	-95.4376	
Eufaula	Site 8	6	OWRB	35.23392	-95.4999	
Eufaula	Site 9	6	OWRB	35.22526	-95.5963	
Eufaula	Surface	4	OWRB	35.45436	-95.6129	
Fort Cobb	Site 1	4	OWRB	35.16215	-98.4568	1,619
Fort Cobb	Site 2	5	OWRB	35.18004	-98.4623	
Fort Cobb	Site 3	5	OWRB	35.18973	-98.4734	
Fort Cobb	Site 4	3	OWRB	35.21948	-98.5095	
Fort Cobb	Site 5	4	OWRB	35.20804	-98.4911	
Fort Cobb	Site 6	1	OWRB	35.21919	-98.4804	
Foss	Site 2	4	OWRB	35.56204	-99.2101	3,561
Foss	Site 3	5	OWRB	35.57753	-99.2352	
Foss	Site 4	5	OWRB	35.60198	-99.2327	
Foss	Site 5	5	OWRB	35.61109	-99.2619	
Foss	Surface	4	OWRB	35.53936	-99.1890	
Grand	Site 10	3	OWRB	36.66528	-94.7703	16,908
Grand	Site 11	3	OWRB	36.65250	-94.7183	
Grand	Site 12	3	OWRB	36.69889	-94.7431	
Grand	Site 13	3	OWRB	36.73333	-94.7739	
Grand	Site 2	2	OWRB	36.51028	-94.9650	
Grand	Site 3	2	OWRB	36.54417	-94.9303	
Grand	Site 4	2	OWRB	36.60361	-94.9039	
Grand	Site 5	3	OWRB	36.56361	-94.8606	
Grand	Site 6	1	OWRB	36.56306	-94.7692	
Grand	Site 7	3	OWRB	36.57194	-94.8331	

Grand	Site 8	3	OWRB	36.62278	-94.8433	
Grand	Site 9	3	OWRB	36.63750	-94.8014	
Grand	Surface	3	OWRB	36.47472	-95.0347	
Grand	12	5	GRDA	36.64998	-94.7084	
Grand	13	6	GRDA	36.68269	-94.7728	
Grand	14	5	GRDA	36.57437	-94.7899	
Grand	15	5	GRDA	36.62237	-94.9079	
Grand	18	1	GRDA	36.49758	-95.0109	
Grand	2	1	GRDA	36.62441	-94.9007	
Grand	21	1	GRDA	36.50138	-94.9242	
Grand	22	6	GRDA	36.55447	-94.8449	
Grand	26	5	GRDA	36.56396	-94.9128	
Grand	27	1	GRDA	36.53917	-94.8389	
Grand	28	1	GRDA	36.53807	-94.8352	
Grand	29	5	GRDA	36.54386	-94.8436	
Grand	7	6	GRDA	36.49768	-94.9185	
Grand	8	1	GRDA	36.50451	-94.9646	
Greenleaf	Site 2	4	OWRB	35.63017	-95.1600	229
Greenleaf	Site 3	4	OWRB	35.64416	-95.1522	
Greenleaf	Site 4	2	OWRB	35.64852	-95.1408	
Greenleaf	Site 5	2	OWRB	35.62183	-95.1605	
Greenleaf	Surface	3	OWRB	35.61719	-95.1664	
Hefner	Site 2	3	OWRB	35.56324	-97.6041	1,000
Hefner	Site 3	3	OWRB	35.56239	-97.5827	
Hefner	Site 4	1	OWRB	35.55528	-97.5929	
Hefner	Site 5	1	OWRB	35.57542	-97.5896	
Hefner	Surface	2	OWRB	35.58102	-97.5974	
Hudson	34	2	GRDA	36.37373	-95.1225	4,856
Hudson	35	2	GRDA	36.30724	-95.1813	
Hudson	87	1	OWRB	36.82300	-96.0476	
Hulah	85	1	OWRB	36.93107	-96.1030	1,445
Jean Neustadt	90	1	OWRB	34.28478	-97.1710	187
Kaw	Site 2	5	OWRB	36.74621	-96.8831	6,879
Kaw	Site 3	4	OWRB	36.76728	-96.8269	
Kaw	Site 4	3	OWRB	36.79969	-96.8291	
Kaw	Site 5	3	OWRB	36.79038	-96.9080	
Kaw	Surface	2	OWRB	36.70118	-96.9242	
Keystone	Site 10	4	OWRB	36.19266	-96.3167	10,523
Keystone	Site 2	3	OWRB	36.19016	-96.2537	
Keystone	Site 3	4	OWRB	36.23228	-96.3001	
Keystone	Site 4	4	OWRB	36.23771	-96.3551	

Keystone	Site 6	4	OWRB	36.16468	-96.2934	
Keystone	Site 8	4	OWRB	36.13969	-96.3285	
Keystone	Site 9	4	OWRB	36.16656	-96.3126	
Keystone	Surface	3	OWRB	36.14726	-96.2573	
Keystone	89	1	OWRB	36.15061	-96.4397	
New Spiro	Site 2	3	OWRB	35.19968	-94.6198	83
Okemah	Site 2	2	OWRB	35.52537	-96.3203	308
Okemah	Site 3	2	OWRB	35.52375	-96.3323	
Okemah	Site 4	1	OWRB	35.50891	-96.3231	
Overholser	Site 2	4	OWRB	35.49974	-97.6766	640
Overholser	Site 3	3	OWRB	35.50848	-97.6700	
Overholser	Site 5	1	OWRB	35.50661	-97.6777	
Overholser	Surface	4	OWRB	35.48716	-97.6682	
Overholser	75	1	OWRB	35.49774	-97.6793	
Perry	Site 2	4	OWRB	36.24243	-97.3409	227
Perry	Site 3	3	OWRB	36.23827	-97.3496	
Perry	Site 5	2	OWRB	36.24680	-97.3398	
Ponca	Site 2	2	OWRB	36.73613	-97.0343	326
Ponca	Site 4	2	OWRB	36.72930	-97.0294	
Ponca	Site 5	2	OWRB	36.71885	-97.0167	
RC Longmire	Site 2	3	OWRB	34.75152	-97.0511	301
RC Longmire	Site 4	2	OWRB	34.75060	-97.0447	
RC Longmire	Site 5	1	OWRB	34.75533	-97.0521	
Rocky (Hobart)	Site 2	2	OWRB	35.17303	-99.0768	138
Rocky (Hobart)	Site 3	3	OWRB	35.18312	-99.0738	
Rocky (Hobart)	Site 4	2	OWRB	35.17694	-99.0788	
Taylor (Marlow)	Site 2	2	OWRB	34.75048	-97.9278	92
Tenkiller Ferry	Site 2	7	OWRB	35.67443	-94.9764	5,221
Tenkiller Ferry	Site 3	5	OWRB	35.73905	-94.9543	
Tenkiller Ferry	Site 4	3	OWRB	35.75542	-94.9051	
Tenkiller Ferry	Site 6	2	OWRB	35.76634	-94.8872	
Tenkiller Ferry	Site 7	6	OWRB	35.63938	-95.0146	
Tenkiller Ferry	Surface	2	OWRB	35.60002	-95.0446	
Tenkiller Ferry	91	1	OWRB	35.75412	-94.9149	
Texoma	Site 10	3	OWRB	33.78593	-96.7962	35,613
Texoma	Site 11	4	OWRB	33.86820	-96.8383	
Texoma	Site 12	4	OWRB	33.89428	-96.8888	
Texoma	Site 2	3	OWRB	33.89302	-96.6113	
Texoma	Site 3	3	OWRB	33.95756	-96.5871	
Texoma	Site 4	3	OWRB	34.00762	-96.6299	
Texoma	Site 6	4	OWRB	33.85604	-96.6915	

Texoma	Site 7	4	OWRB	33.82814	-96.7376	
Texoma	Site 8	4	OWRB	33.84694	-96.7801	
Texoma	Site 9	4	OWRB	33.81989	-96.8054	
Texoma	Surface	4	OWRB	33.82997	-96.5768	
Texoma	Buncombe North 5	1	PELL	33.87486	-96.8073	
Texoma	Buncombe Pelagic 2	30	PELL	33.87125	-96.8074	
Texoma	Buncombe South 1	1	PELL	33.86158	-96.8069	
Texoma	Dam North	1	PELL	33.83689	-96.5892	
Texoma	Dam Pelagic 4	27	PELL	33.82319	-96.5902	
Texoma	Islands Pelagic 3	30	PELL	33.82875	-96.7304	
Texoma	Red River Pelagic 1	30	PELL	33.89742	-96.8874	
Texoma	Washita Pelagic 5	29	PELL	33.96403	-96.5769	
Texoma	Washita South 1	1	PELL	33.89358	-96.5808	
Thunderbird	1	42	OWRB	35.22333	-97.2208	2,456
Thunderbird	2	56	OWRB	35.23889	-97.2289	
Thunderbird	3	54	OWRB	35.26222	-97.2389	
Thunderbird	4	56	OWRB	35.22444	-97.2508	
Thunderbird	5	54	OWRB	35.22028	-97.2906	
Thunderbird	7	23	OWRB	35.20306	-97.2581	
Thunderbird	8	31	OWRB	35.28641	-97.2449	
Thunderbird	Fisherman's point	4	PELL	35.22862	-97.2460	
Thunderbird	North Sentinel	4	PELL	35.23219	-97.3075	
Webbers Falls	Site 2	4	OWRB	35.60167	-95.1817	4,694
Webbers Falls	Site 4	2	OWRB	35.63056	-95.2717	
Webbers Falls	Site 6	4	OWRB	35.70000	-95.2314	
Webbers Falls	Surface	3	OWRB	35.55472	-95.1706	
WRHoloway	Site 2	6	OWRB	36.24564	-95.0996	318
WRHoloway	Site 3	9	OWRB	36.25695	-95.0868	
WRHoloway	Site 4	3	OWRB	36.25100	-95.1021	
WRHoloway	Site 5	1	OWRB	36.25209	-95.0931	
WRHoloway	Surface	10	OWRB	36.25634	-95.1023	
WRHoloway	36	2	GRDA	36.25576	-95.1006	

WRHoloway	38	1	GRDA	36.25562	-95.0900	
-----------	----	---	------	----------	----------	--

Table S2. Site information for the satellite ground-truthed phycocyanin samples. Sampling organization included the Plankton Ecology and Limnology Lab (PELL).

Lake	Site	Number of samples	Sampling organization	Latitude	Longitude
Texoma	Buncombe North 5	1	PELL	33.87486	-96.8073
Texoma	Buncombe Pelagic 2	17	PELL	33.87125	-96.8074
Texoma	Buncombe South 1	1	PELL	33.86158	-96.8069
Texoma	Dam North	1	PELL	33.83689	-96.5892
Texoma	Dam Pelagic 4	17	PELL	33.82319	-96.5902
Texoma	Islands Pelagic 3	17	PELL	33.82875	-96.7304
Texoma	Red River Pelagic 1	17	PELL	33.89742	-96.8874
Texoma	Washita Pelagic 5	17	PELL	33.96403	-96.5769
Texoma	Washita South 1	1	PELL	33.89358	-96.5808
Thunderbird	Fisherman's Point	4	PELL	35.22862	-97.2460
Thunderbird	North Sentinel	4	PELL	35.23219	-97.3075

Table S3. Ground-based site information. Sampling organization included the Plankton Ecology and Limnology Lab (PELL).

Lake	Site	Number of samples	Sampling organization	Latitude	Longitude
Ellsworth	Dam	1	PELL	34.79550	-98.36640
El Reno	1	1	PELL	35.52718	-97.98684
El Reno	2	1	PELL	35.52373	-97.98925
El Reno	3	1	PELL	35.51960	-97.99631
El Reno	4	1	PELL	35.52640	-97.98934
El Reno	5	1	PELL	35.52110	-97.99271
Grand	Drip	4	PELL	36.49969	-94.95614
Grand	Drown	4	PELL	36.49842	-94.91957
Grand	Duck	5	PELL	36.53644	-94.97219
Grand	Grand	5	PELL	36.68269	-94.77286
Grand	Honey	5	PELL	36.57511	-94.78772
Grand	Horse	5	PELL	36.62122	-94.90856
Grand	IS1	5	PELL	36.49250	-95.04489
Grand	IS3	5	PELL	36.50858	-94.95539
Grand	Sail	5	PELL	36.64175	-94.81478
Grand	Tree	5	PELL	36.56339	-94.91283
Grand	Wood	5	PELL	36.53644	-94.82236
Overholser	1	1	PELL	35.48720	-97.66820
Overholser	2	1	PELL	35.49970	-97.67660
Overholser	4	1	PELL	35.49230	-97.67560
Rocky	Dam	1	PELL	35.16720	-99.07430
Texoma	Buncombe North 4	1	PELL	33.87317	-96.80725
Texoma	Buncombe North 5	3	PELL	33.87472	-96.80722
Texoma	Buncombe Pelagic 2	4	PELL	33.87125	-96.80744
Texoma	Buncombe South 1	2	PELL	33.86158	-96.80692
Texoma	Buncombe South 2	1	PELL	33.86572	-96.80764
Texoma	Dam North	3	PELL	33.83689	-96.58919
Texoma	Dam Pelagic 4	4	PELL	33.82319	-96.59022
Texoma	Islands Pelagic 3	4	PELL	33.82875	-96.73044
Texoma	Red River Pelagic 1	3	PELL	33.89481	-96.89172

Texoma	Washita Pelagic 5	4	PELL	33.96403	-96.57692
Texoma	Washita South 1	3	PELL	33.87553	-96.61975
Texoma	Washita South 2	3	PELL	33.92900	-96.57131
Thunderbird	1	2	PELL	35.22275	-97.22237
Thunderbird	2	2	PELL	35.23855	-97.22916
Thunderbird	3	2	PELL	35.26260	-97.23883
Thunderbird	4	2	PELL	35.22381	-97.25139
Thunderbird	5	1	PELL	35.22016	-97.28863
Thunderbird	6	1	PELL	35.23065	-97.30523
Thunderbird	8	2	PELL	35.28664	-97.24446
Thunderbird	11	2	PELL	35.21296	-97.30335
Thunderbird	13	1	PELL	-	-
Thunderbird	Fisherman's Point	4	PELL	35.22909	-97.24615
Thunderbird	North Sentinel	4	PELL	35.23194	-97.30929
Thunderbird	Sailboat	4	PELL	35.23037	-97.23629

Table S4. Performance of satellite algorithms on the resampled data set (N = 511) for chlorophyll-*a* estimation on Oklahoma lakes was evaluated using the adjusted R² from the given linear model, the root mean square error (RMSE), and Pearson's r correlation coefficient. *The RMSE for the resampled model should not be directly compared to the other models because the dataset used to build the model was different

Algorithm name		Adjusted R ²	Model P-value	RMSE	Pearson's r	Pearson's r P-value
Chlorophyll- <i>a</i> algorithms						
1	NDCI (Normalized Difference Chlorophyll Index)	-0.0013	0.57	22.26	0.025	0.57
2	SABI (Surface Algal Bloom Index)	-0.0014	0.58	22.26	0.024	0.58
3	FLHB (Fluorescence Line Height algorithm blue)	-0.0014	0.59	22.26	0.024	0.59
4	2BDA (two-band algorithm)	-0.0007	0.42	22.26	0.036	0.42
5	3BDA (three-band algorithm)	-0.0007	0.43	22.26	0.035	0.43
6	KIVU (3BDA – like)	-0.0009	0.46	22.26	0.029	0.46
7	this paper (resampled dataset)	0.025	<0.001	21.95	0.170	<0.001

Equation for the ground-based chlorophyll-*a* model. For band information see Table 2.

$$\begin{aligned} \text{CHLA} = & -222.71 - 100.43(\text{Band1:Band3}) + 107.2(\text{Band1:Band5}) \\ & - 34.52(\text{Band2:Band3}) + 564.61(\text{Band3:Band4}) - 856.16(\text{Band3:Band5}) \\ & + 564.63(\text{Band4:Band5}) \end{aligned} \quad (\text{eq. S1})$$

Equation for the ground-based phycocyanin model. For band information see Table 2.

$$\begin{aligned} \text{PCY} = & -1311.71 - 161.24 (\text{Band1:Band2}) + 145.35 (\text{Band1:Band5}) + \\ & 568.31 (\text{Band2:Band3}) - 836.64 (\text{Band2:Band4}) + 2074.41 (\text{Band3:Band4}) \\ & - 1370.80 (\text{Band3:Band5}) + 948.89 (\text{Band4:Band5}) \end{aligned} \quad (\text{eq. S2})$$

R code for processing Landsat images in the R environment:

```
require(raster)
require(rgdal)
require(dplyr)
require(sp)

#set your working directory, all of your image folders should be in this folder
your_working_directory <- "K:/Images for qaqc - 10Nov20/Thunderbird8"

setwd(your_working_directory)

file.names <- list.files(your_working_directory)

for (x in 1:length(file.names)){
  print(x)
  holder <- paste0(your_working_directory, file.names[x], "/")
  file.names[x] <- holder
}

field_points <- read.csv(file="K:/Images for qaqc - 10Nov20/Thunderbird8.csv") #csv has to be
set up with longitude in a column before latitude, can have as many lake site combinations as
needed

#field_points2 <- select(field_points, -Lake)
```

```
field_points2 <- project(as.matrix(field_points[3:4]), proj="+proj=utm +zone=14
ellps=WGS84") ##use this if you need to change the projection of the field points, all of the
images should be in the same projection for each field point file ##for field_points[x:y] put the
longitude (x) and latitude (y) columns
```

```
#calculate the extent of the image you need to look at based on field points
e <- extent(min(field_points2[,1])-1000, max(field_points2[,1])+1000,
            min(field_points2[,2])-1000,max(field_points2[,2])+1000)
```

```
for(j in 1:length(file.names)){
```

```
  bands = list.files(path = paste(file.names[j]), pattern = "sr_band[1-7].tif$", recursive = TRUE)
  ##make sure this is correct
```

```
  for(i in 1:6){
    r=raster(paste(paste(file.names[j]),bands[i],sep=""))
    r <- crop(r, e)
    r <- focal(r, w=matrix(1/9, nrow=3, ncol=3), na.rm=TRUE)
    names(r) <- paste("layer",i)
    if(i==1){
      b=r
    }
    else {
      b=addLayer(b,r)
    }
    print(i)
  }
}
```

```
#extract points of interest (these represent the average of the 3x3 neighborhood of the point of
interest)
```

```
tempDF3x3 <- as.data.frame(extract(b,field_points2))
```

```
#add info about site
```

```
tempDF3x3$site <- c(seq(1,1))
```

```
#add filename to df
```

```
tempDF3x3$filename <- c(rep(file.names[j],1)) #is this the number of sites?
```

```
#rename columns that hold the band data
```

```
names(tempDF3x3)[1:6] <- c("band_1", "band_2", "band_3", "band_4", "band_5", "band_7")
```

```
if(j==1){
  masterDF <- tempDF3x3
}
else{
```

```
    masterDF <- bind_rows(masterDF, tempDF3x3)
  }
}
```

```
#save the surface reflectance to a csv
write.csv(masterDF, file='./Landsat_sr.csv')
```