



# Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs

Michael Papenfus · Blake Schaeffer · Amina I. Pollard · Keith Loftin

Received: 15 October 2019 / Accepted: 24 September 2020 / Published online: 2 December 2020  
© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2020

**Abstract** Assessment of chlorophyll-a, an algal pigment, typically measured by field and laboratory in situ analyses, is used to estimate algal abundance and trophic status in lakes and reservoirs. In situ-based monitoring programs can be expensive, may not be spatially, and temporally comprehensive and results may not be available in the timeframe needed to make some management decisions, but can be more accurate, precise, and specific than remotely sensed measures. Satellite remotely sensed chlorophyll-a offers the potential for more geographically and temporally dense data collection to support estimates when used to augment or substitute for in situ measures. In this study, we compare available chlorophyll-a data from in situ and satellite

imagery measures at the national scale and perform a cost analysis of these different monitoring approaches. The annual potential avoided costs associated with increasing the availability of remotely sensed chlorophyll-a values were estimated to range between \$5.7 and \$316 million depending upon the satellite program used and the timeframe considered. We also compared sociodemographic characteristics of the regions (both public and private lands) covered by both remote sensing and in situ data to check for any systematic differences across areas that have monitoring data. This analysis underscores the importance of continued support for both field-based in situ monitoring and satellite sensor programs that provide complementary information to water quality managers, given increased challenges associated with eutrophication, nuisance, and harmful algal bloom events.

M. Papenfus ()  
Office of Research & Development, U.S. Environmental Protection Agency, Corvallis, OR 97330, USA  
e-mail: papenfus.michael@epa.gov

B. Schaeffer  
Office of Research & Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA  
e-mail: schaeffer.blake@epa.gov

A. I. Pollard  
Office of Water, U.S. Environmental Protection Agency, Washington, DC 20460, USA  
e-mail: pollard.amina@epa.gov

K. Loftin  
U.S. Geological Survey, Kansas Water Science Center, Lawrence, KS 66049, USA  
e-mail: kloftin@usgs.gov

**Keywords** Chlorophyll · Harmful algal blooms · Water quality monitoring · Remote sensing · Lakes · Economic value

## Introduction

Chlorophyll-a as an indicator of eutrophication and harmful algal blooms

Ecosystem stress in lakes and reservoirs has historically been evaluated based on water quality condition and biological integrity using a suite of laboratory and field

(*in situ*) measures. Algae are a critical component of lake food webs and support primary consumers. Owing to its basal position in lake food webs, algae is responsive both to changes among higher trophic levels, e.g., zooplankton and fish, and to lake chemical characteristics. Eutrophication caused by excess nutrients is one type of ecosystem stress that can lead to the formation of an overabundance of algae and cyanobacteria, both of which can have negative environmental or health consequences and are collectively known as harmful algal blooms (HABs) (Codd 2000; Smayda 1997). While freshwater HAB events may be associated primarily with cyanobacteria, there are only one HAB freshwater taxa. Other taxa include Haptophytes, Euglenophytes, Raphidophytes, and Dinoflagellates that all may produce ichthyotoxins. Chlorophyte, cryptophyte, and diatom blooms cause hypoxia and other negative environmental consequences. Chlorophyll is representative of all these taxa, and therefore for the purpose of this study, we do not limit the scope to only cyanobacterial HABs (cyanoHAB) but include the broader HAB taxa.

A range of environmental conditions can render algal blooms harmful (e.g., low dissolved oxygen, nuisance taste-and-odor compounds, disinfection by-product formation in drinking water treatment) leading to undesired ecosystem and socioeconomic consequences. CyanoHABs, one of many groups of freshwater algae that may cause harm, are frequently encountered and can sometimes produce cyanotoxins that present a formidable, ecological, and human health challenge that may be the most difficult outcome to predict (Backer et al. 2013; Graham et al. 2009; Loftin et al. 2016; Stewart et al. 2006). Human health protection from cyanobacteria exposure can require rapid risk management decisions often with insufficient data while attempting to minimize unnecessary socioeconomic consequences. The drinking water advisory issued for Salem, Oregon, in 2018, is a good example of such an event (City of Salem 2018). Remote sensing may be able to fill in data gaps left by *in situ*-dependent approaches.

All phytoplankton, including cyanobacteria, are photosynthetic primary producers and therefore produce chlorophyll-a (chl-a) as well as other accessory pigments (Stumpf et al. 2016). As algal populations increase, chl-a concentrations increase. Taking advantage of this relation, chl-a concentration is frequently used as a surrogate for phytoplankton bloom concentration and may serve as an indicator for excess nutrients (Devlin

et al. 2011; Ferreira et al. 2011; Schaeffer et al. 2011). Eutrophic and hyper-eutrophic lakes are more likely to experience HAB events and nuisance algae than nutrient poor waterbodies, and there is growing concern about these nutrient-driven events across the USA and elsewhere (Heisler et al. 2008; Smith and Schindler 2009). The 2012 National Lakes Assessment (NLA) estimated that 55% of lakes in the continental USA (CONUS) are either eutrophic or hyper-eutrophic based on chl-a measures (U.S. Environmental Protection Agency [USEPA] 2016). These events have the potential to cause economic damage through impacts to human, animal, and ecological health (Ho and Michalak 2015). They also diminish a variety of socioeconomic benefits through reduced recreational opportunities and reduced aesthetics, cause taste and odor problems in drinking water supplies, and cause harm to aquatic ecosystems through a variety of pathways (Brooks et al. 2016). There are special circumstances where chl-a can be an indicator for specific groups of algae such as with cyanobacteria when they are the main population of algae present; typically chl-a alone is not the best indicator for a specific group of phytoplankton without further measurements such as accessory pigments specific to the groups of algae present (Loftin et al. 2016; Stumpf et al. 2003; Stumpf et al. 2016).

The contributions of this study include the following: (a) a comparison of the spatial and temporal distribution of available chl-a observations across different *in situ* and satellite-based datasets on a national scale; (b) computation of the avoided costs associated with making remotely sensed data available nationally in lieu of *in situ* sampling; and (c) characterization of land ownership and select sociodemographic variables surrounding waterbodies that are resolvable in both types of datasets. In this application, avoided costs are a measure of what it would cost to “replace” the satellite observations using data obtained from *in situ* methods on the same spatial and temporal scale as is available using either land or ocean satellite datasets assuming that the information and therefore the value of chl-a measured by both approaches is equivalent. There may be situations depending on how the data are used where this is not a valid assumption due to differences in precision, accuracy, or specificity (e.g., thresholds, mechanistic modeling, characterization of chl-a degradation products called pheopigments). This manuscript does not portend to resolve all of these nuances, but rather puts forward an economic method for evaluating these differences to

better understand the value of remotely sensed measures versus in situ measures of chl-a to help the end-user better utilize both approaches as needed to meet objectives. The avoided cost is the difference in costs between in situ sampling costs and the costs associated with providing the satellite observations for use in applied settings.

#### In situ water quality monitoring and data availability

Despite widespread eutrophication and associated nuisance or HAB events, there remains limited, long-term, coordinated, and consistent monitoring programs that collect data to monitor trends in bloom events. This monitoring data is also essential for evaluating the causes, consequences, and management approaches required to reduce the occurrence of these events (Brooks et al. 2016; Smith et al. 2014). Strategies to support the attainment of safe and healthy water quality in freshwater bodies include in situ water quality management programs. Most in situ water quality monitoring efforts in the USA are focused on assessing and addressing local and regional water quality issues where more than 90% of waterbodies contained less than 5 stations, and more than 55% only had a single monitoring station (Schaeffer et al. 2018a).

The Water Quality Portal (WQP) was developed as a publicly accessible database to simplify dissemination of water quality data (U.S. Geological Survey [USGS], USEPA, and National Water Quality Monitoring Council 2017). In addition to voluntary state data uploads, the WQP contains data from the USEPA STOrage and RETrieval Water Quality Exchange (STORET) (USEPA 2017), the USGS National Water Information System (NWIS) (USGS 2017a)(USGS 2017a), and the U.S. Department of Agriculture (USDA) Sustaining the Earth's Watersheds Agricultural Research Data System, or STEWARDS (2017). The resulting WQP database contains data from both state and federal programs. Although the WQP currently provides > 3 million water quality records from all 50 states, the data in the WQP do not conform to consistent sample collection, laboratory processing, analytical methods, or reporting protocols (Read et al. 2017). The WQP includes a wide variety of biophysical and chemical characteristics of the monitored waterbodies, including typical water quality measures such as nutrient concentrations, fecal coliform bacteria abundance, total suspended solid concentration, and temperature, in addition to chl-a

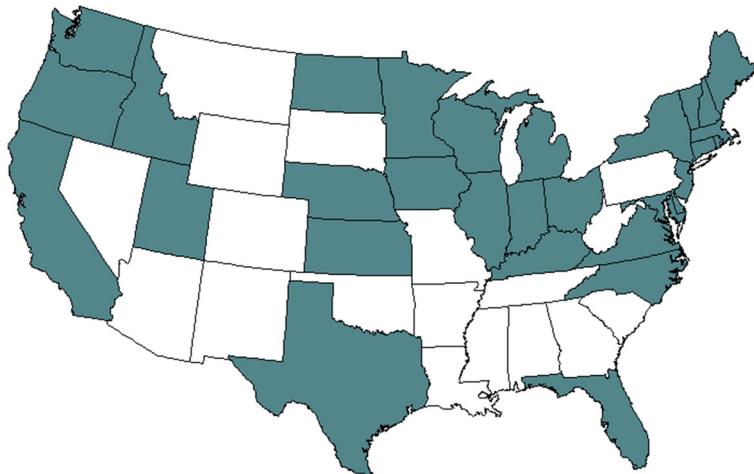
concentrations. The data in the WQP may or may not be associated with HAB-specific monitoring programs and state contributions to WQP are more expansive than states with current HAB programs (see Fig. 1). The WQP data are included in this comparative analysis because they represent a compiled, publicly accessible set of in situ water quality measures across the USA at the local and state level.

The NLA survey is a part of the USEPA effort to work collaboratively with states and tribes to assess the condition of lakes, ponds, and reservoirs across the USA (USEPA 2016). The NLA collects coordinated, consistent field data to develop broad-scale conclusions about the biological, chemical, physical, and recreational condition of surface waters across the USA. The sampled waterbodies are selected from a probability-based sample design that allows for population-level inferences to the larger set of 110,000 lakes included in the sampling frame (Pollard et al. 2018). As a part of the assessment of recreational lakes, NLA includes a measure of microcystins as well as nuisance and HAB biomass indicators such as cyanobacteria abundance and chl-a in lakes. While the NLA includes these HAB indicators, it was not designed as a comprehensive HAB monitoring program. It does provide a unique perspective on observations from across a wide range of lakes and reservoirs across the USA and is readily available from public websites.

#### Satellite derived chlorophyll-a

Previous studies provide a comprehensive review of past, present, and new satellite sensors available for deriving water quality in lakes, reservoirs, and other inland waters (Dörnhöfer and Oppelt 2016; Tyler et al. 2016). The advantages and disadvantages of different sensors with varying spatial, spectral, and temporal resolutions, along with recent progress updates are discussed in other studies (Mouw et al. 2015; Palmer et al. 2015; Greb et al. 2018). Lake and reservoir chl-a is detectable from satellite remote sensing (Gitelson 1992; Wynne et al. 2008; Wynne et al. 2010). There are numerous examples of the effectiveness of using satellite remotely sensed data to detect HAB biomass (Anderson 2009; Kutser et al. 2006; Dekker et al. 2018), but little information exists on the suitability of these satellite sensors for state and national monitoring programs in terms of their spatial and temporal coverage when compared with traditional field-based programs (Clark et al. 2017).

**Fig. 1** States with HAB monitoring programs and information. Source: <https://www.epa.gov/cyanohabs/state-habs-monitoring-programs-and-resources>



### Satellite remote sensing as a complement to traditional in situ monitoring

Different monitoring approaches provide unique scales of information relevant to managing and understanding HAB events because of the temporal, spatial, and compositional variability of HAB and nuisance algae. Traditional field-based sampling is incorporated into most existing water quality monitoring programs. This field sampling provides measured information for numerous biological and chemical water variables (e.g., nutrient concentrations), and is currently the only approach that can confirm the presence of toxic compounds in water. However, traditional field-based monitoring programs can be expensive in terms of labor, travel, and equipment. These programs may not be spatially and temporally representative of ambient water quality conditions and results may not be available in a timeframe that is relevant to protecting human health and safety (Schaeffer et al. 2013). More frequent sampling over time and space based on satellite remote sensing has the potential to improve the representativeness of data that can be provided to water quality managers. Remotely sensed data can also improve in situ sampling efficiency through targeted allocation of resources and subsequent cost savings. However, the caveats to using satellites include a limited number of derived water quality variables; issues with the accuracy and precision of measurements; specificity (e.g., pheophytin, accessory pigments with overlapping absorption spectra) may be less than what is encountered in field and laboratory measurements; and the inability to measure the concentration of toxins caused by HABs. The spatial resolution that generally excludes small waterbodies and nearshore areas and cloud cover may limit the temporal

frequency of usable images throughout the year. Remote sensing approaches also require a different set of technical expertise to process and interpret the imagery, although some steps can be automated. Perhaps most importantly, algorithms to detect and measure water quality constituents from remote sensing data require in situ measures for algorithm development, calibration, and validation.

All measurements are subject to error. It is a misnomer to assume in situ measures contain no error, are a “ground-truth,” or a better representation of the environment any more than satellite-derived water quality measures are devoid of error. Models, satellite algorithms, and in situ measures are all approximate representations of the environment (Wainwright and Mulligan 2013) where fitness for purpose would need to be evaluated for the selected approach. In situ measures may contain significant error depending on the analysis method, where standard fluorometric methods generally underestimate chl-a concentrations compared with high-performance liquid chromatographic methods (Trees et al. 1985). The North American Lake Management Society (NALMS n.d.; <https://www.nalms.org/secchidipin/monitoring-methods/chlorophyll-analysis/>) also clarifies that there is little evidence that results derived by different methods are similar. In addition to in situ measurement error, there can be significant spatial difference between using single point in situ samples versus depth-width integrated in situ sampling approaches and an integrated detection of a satellite pixel over a spatial area and depth of the water column. However, single point in situ sampling is quite common in lakes and reservoirs. Therefore, in situ and satellite-derived chl-a would rarely fit a perfect one-to-one regression line. There is still valuable information in the satellite and in situ measures even if the chl-a are not

exactly one-to-one. For example, from a HAB perspective, it has been confirmed that just presence and absence information is useful in issuing recreational advisories (Schaeffer et al. 2018b). The next level of quantification would be categorical, where chl-a is classified into trophic categories. The final level of quantification is concentration values, but in situ sampling could likely be improved if depth-width integrated sampling would be utilized. Seegers et al. (2018) clarified that when considering satellite algorithm performance, decisions would benefit from not being limited to just accuracy. Instead users could consider the impact of the dynamic range of the data, algorithm stability, spatial coverage, uncertainty, and bias. For example, trend data may be better characterized by bias and consistency over accuracy. Satellite remote sensing of water quality methods will continue to mature over the coming decade (Greb et al. 2018) as new approaches continue to be developed such as demonstrated by Pahlevan et al. (2020).

Remote sensing and field-based in situ monitoring approaches offer different, but complementary information. Consequently, there is an opportunity and need to explore the efficiencies and potential cost savings associated with a combined approach.

Some US states have established programs for monitoring nuisance or HAB events (Fig. 1). These programs vary across states with most efforts being focused on event response, but some are more systematic and long-term in nature. For instance, California has established a statewide, long-term strategic plan to monitor and respond to HAB events, assess risk for future HAB events, and provide public education and outreach. One of the monitoring options in this strategy is the adoption of satellite imagery to identify and track HABs (Anderson-Abbs et al. 2016). The California strategic plan uses satellite imagery to conduct historical analysis of HABs to assess waterbodies at risk and to prioritize field monitoring, remediation, and future management. California also uses imagery to develop protocols for alerting water quality managers and communities as to when and where cyanobacteria blooms are occurring. In addition to the use of satellite imagery, California's plan includes the integration of this data with existing water quality, watershed, and volunteer field-based assessment and ambient monitoring programs. However, existing field-based monitoring programs in California are limited by resource constraints and the fact that very few freshwater lakes with substantial recreation activity and drinking water use are monitored regularly within the state (Anderson-Abbs et al. 2016).

Ohio uses a combination of methods to confirm that cyanobacteria are present including satellite imagery, microscopic examination, and additional screening tools (State of Ohio 2016). Other states like New York also have extensive monitoring and public information systems that provide weekly updates on monitored waterbodies and current HAB conditions. Most states currently rely on visual observations from field staff or citizens. These reports result in field visits to collect samples and laboratory analysis to determine whether a waterbody contains cyanobacteria (as well as sometimes measuring the concentrations of toxins) or other nuisance species (Graham et al. 2009).

## Methods

This study was limited to free and publicly available data including land and ocean satellite missions with spatial resolutions most appropriate for lakes and reservoirs. The satellite data examined in this study were derived from the Landsat-8 Operational Land Imager (OLI) mission (L8) and the European Space Agency (ESA) Copernicus Sentinel-3 Ocean and Land Colour Imager (OLCI) sensor (S3). Both L8 and S3 data provide new and improved opportunities for monitoring inland water quality because of their potential to detect chl-a. However, it is important to note that operational delivery of chl-a data for US lake and reservoir water quality management is mostly theoretical and has only recently been demonstrated in a limited capacity, such as through the Cyanobacteria Assessment Network (CyAN) using S3 (Schaeffer et al. 2015). There has been limited demonstration of satellite-derived products specific for aquatic environments at the US scale using L8 or other land imagers, such as Secchi depth measures for the Upper Midwestern USA (Chipman et al. 2004), surface temperatures (Schaeffer et al. 2018a; <https://www.usgs.gov/land-resources/nli/landsat/landsat-surface-temperature>), and provisional aquatic reflectance (Franz et al. 2015; Pahlevan et al. 2017; <https://www.usgs.gov/land-resources/nli/landsat/landsat-provisional-aquatic-reflectance>). S3 is well-suited for monitoring chl-a because of the required spectral bands and frequent 2- to 3-day revisit times, where L8 has a 16-day revisit time, higher spatial resolution, but limited spectral band options. The ESA Sentinel-2 Multispectral Imager (MSI) may also provide measures on similar spatial and

temporal scales to that of L8 (Toming et al. 2016), so results for L8 are assumed to be representative here.

### Spatial and temporal sampling

In situ discrete measures of chl-a were accessed through the WQP (USGS, USEPA, and National Water Quality Monitoring Council 2017) and the 2012 NLA (USEPA 2016). The data selected from the WQP consisted of chl-a surface water samples collected from lakes, reservoirs, or impoundment sites from 1/1/1980 through 1/1/2016 within CONUS.

Inland lakes and reservoirs with discrete in situ chl-a samples from both the 2012 NLA and the WQP were compared against L8 and S3 resolvable waterbodies. Landsat path/row World Reference System polygon shapefiles were acquired through the USGS Landsat acquisition tool (USGS 2017b), and S3 swath polygon shapefiles were acquired through ESA Earth Observation Swath and Orbit Visualization software (ESOV 2017). Resolvable waterbodies were previously calculated from Clark et al. (2017); those that fit the NLA 2012 site evaluation criteria and the method are briefly summarized here.

A waterbody mask was generated using the NHD Plus version 2.0 (McKay et al. 2012) to identify waterbodies resolved with 300-m or 30-m pixel resolution, assuming a minimum three-by-three-pixel array requirement. The waterbody spatial coverage with resolvable satellite pixels was calculated based on the minimum Euclidian distance from shore that will accommodate the three-by-three-pixel or larger array. This resulted in 1862 resolvable waterbodies by S3 and 170,240 resolvable waterbodies by L8. Cloud-free views were calculated using the Terra MODIS 5-km daytime cloud mask dataset covering the period 2001–2010 (Mercury et al. 2012). The mean cloud-free percentage was calculated from all raster cell values intersecting the waterbody polygon<sup>1</sup>.

<sup>1</sup> It should be noted that the supplemental data files provided by the Mercury et al. (2012) journal article have been compressed and altered by the hosting journal service, so they were not usable as provided through the publication. We had to obtain the original GeoTIFFS directly from the authors. Here we use GeoTIFF values, where 0 is equal to 100% cloud-free and 255 is equal to 0% views being cloud-free on a linear scale as confirmed by the co-authors. The equation used to convert from GeoTIFF digital number to cloud-free observations is reversed from that suggested in the manuscript, because the files provided for our analysis were percent cloudiness, not cloud-free views.

A waterbody was considered viewed once by the satellite platform if the entire waterbody polygon was contained within the swath path. The number of satellite swath path views was multiplied by mean cloud-free percentage and number of revisits per year. For example, the L8 satellite has a 16-day revisit time, where each path and row are passed over 22–23 times in a year. A lake completely contained within a single L8 swath with 35% cloud-free observations has the potential for 8 observations per year.

### Avoided costs

In addition to identifying waterbodies across the USA that include in situ chl-a observations or those resolvable by satellite for chl-a estimation, we also provide a simplified analysis of the avoided costs associated with adopting satellite-derived estimates of chl-a. The avoided cost in this analysis is narrowly defined as the difference between the laboratory cost of analyzing in situ water samples for chl-a and the costs associated with providing the satellite-derived chl-a observations for use in applied settings. We motivate the analysis of avoided costs by asking the following question: if satellite-derived chl-a observations were not available, how much would it cost using in situ methods to produce observations on the same spatial and temporal scale as is potentially available from satellite observations? This is particularly relevant for water quality monitoring programs that are focused on HAB and nuisance algae events, which will require more frequently collected spatial, and temporal data. For this analysis, we only consider a portion of the costs associated with obtaining chl-a observations using in situ methods and compare these costs to the expenses associated with making satellite-derived chl-a data available to the public and other entities on a cloud-based data platform.

The conceptual diagram in Fig. 2 illustrates the major steps required to produce useful chl-a observations and to highlight the specific costs considered for this comparison. The steps associated with making in situ chl-a observations available for end user interpretation are field sampling, development of standardized laboratory procedures, and laboratory analysis. Because the cost of field sampling and development of standardized laboratory procedures is highly variable depending on the location and overall purpose of the collection, we do not include these costs in our primary results. However, we do include an example of including travel costs for a

single state to illustrate how these costs could be included in an analysis tailored to a specific region. Figure 2 also illustrates the parallel steps required to produce the satellite-derived observations. This would entail launching a satellite mission, development of algorithms to convert satellite data into usable chl-a measures, and then storing, managing, and making these large datasets available to the public in a usable format. For comparison with the benefits associated with adopting satellite observations, we only compute the expenditures required to store, manage, and make publicly available the satellite-derived data. We do not consider the costs associated with launching a satellite mission or the costs associated with developing algorithms to translate raw satellite data into water quality metrics.<sup>2</sup>

To calculate an approximation of the avoided costs associated with using satellite observations in lieu of in situ sampling to produce data on the same spatial and temporal scale as is available using satellite observations, we use Monte-Carlo error propagation (or uncertainty analysis) to compute a range of potential avoided costs (Lee 2014; Morgan et al. 1990). Because there will be variability in the cost parameters highlighted in Fig. 2, we assign each of the parameters a distribution of possible values and that variability is then propagated through the avoided cost calculations. For this analysis, we assign each variable to a triangular distribution that is defined by the mode and an upper and lower bound. For example, while we provide an annual estimate of the number of daily cloud-free images available for each of the satellite-resolvable lakes, this number will vary from year to year. For our calculations, we set the upper bound of annual observations to the number of cloud-free observations we compute using the data from Mercury et al. (2012) and then conservatively set the mode and lower bound to 85, and 75% of the upper bound for this illustrative example. The annual avoided cost of the use of satellite data as opposed to in situ data is calculated as:

<sup>2</sup> We do not explicitly include the costs associated with launching a satellite mission or the costs associated with developing the algorithms required to translate the raw satellite sensor data into useable water quality metrics. The information used from satellite missions is used in many other applications, so it is not generally feasible to attribute a portion of those costs to a task such as monitoring for chl-a or temperature. Similarly, the costs of algorithm research and development are difficult to measure. For these reasons, we also did not include highly variable costs associated with the field work required to collect the in situ water quality samples. That said, one could modify these analyses to include some representative numbers for each of these tasks.

$$\text{Annual Avoided Cost} = \sum_{i=1}^N \text{Observations}_i \times \text{Sample Cost}_{\text{chl-a}} - \text{Annual Satellite Cost}$$

where the annual cloud-free observations from the satellite are summed across all  $N$  resolvable waterbodies. The sample laboratory analysis costs are intended to be broadly representative of a mean cost per chl-a sample and would likely vary across laboratories and monitoring programs. We compute the present value (PV) of these avoided costs for a 10-year monitoring project. A 10-year project's stream of costs is converted to present values such that:

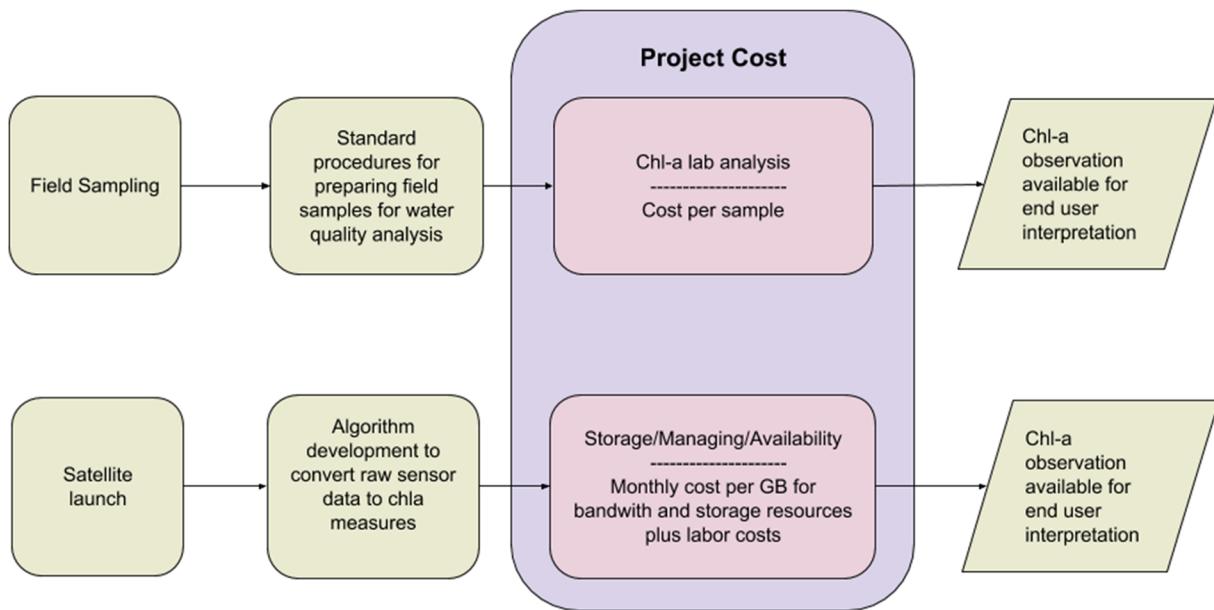
$$PV = AC_0 + d_1 \times AC_1 + d_2 \times AC_2 + \dots + d_{n-1} \times AC_{n-1} + d_n \times AC_n$$

where  $AC$  is the annual avoided cost and the discounting factors are given by  $d_t = \frac{1}{(1+r)}^t$ . We illustrate the present value of the avoided cost using a discount rate,  $r$ , of 3 and 7%<sup>3</sup>.

Table 2 shows each of the variables included in the avoided cost calculations and the parameters used to describe their distributions. The primary costs associated with making the satellite data available to the public and other institutions for monitoring and decision support are data storage/hosting, maintaining and updating the data repository, and making the data available via a cloud-based service as shown in Fig. 2. Commercial cloud service providers base their pricing on storage, computation, and bandwidth (Rajakumari et al. 2014; Yuan et al. 2016). The S3 CONUS data used to derive chl-a estimates consist of 7-day composite images. For this analysis, we assume that server storage for these images is about 200 Gb for staging the individual tiff files based on CyAN data production. Schaeffer et al. (2018b) provide an alternative example in which images are made available for use in a mobile application. The range of storage costs per GB data used in this analysis is based on calculations obtained from Amazon for their Amazon Web Service storage and CloudFront content delivery<sup>4</sup> and may vary depending on how often the data are accessed and transferred. These costs although relatively inexpensive are likely to decrease even further in

<sup>3</sup> USEPA 2014, recommends using a single central rate of 5% as intermediate between 3 and 7% rates. We use these as our lower and upper bounds.

<sup>4</sup> <https://calculator.s3.amazonaws.com/index.html>. (Accessed 03 June 2020).



**Fig. 2** Conceptual diagram of costs explicitly considered

the future (Krumm and Hoffman 2020). A relevant example of declining hosting costs has already been demonstrated with Google Earth Engine hosting S3 and L8 data, where algorithms can be implemented on a free platform, and users can publicly publish their results (Ho and Michalak 2015; Gomarasca et al. 2019). While we provide cost estimates for satellite data hosting, these may be substantially lower going into the future. In addition to the annual costs associated with storing and retrieving the satellite-derived data, we also include the cost associated with providing two federal GS-13 level scientists to provide technical support. This support would include data quality assurance tasks, algorithm updates, and technical support associated with translating the scientific data into meaningful metrics that can be understood by the public and water quality managers. The initial development and maintenance of a user interface is generally leveraged against existing infrastructure for a single product such as chlorophyll. The existing infrastructure is typically used in many other applications such as demonstrated by (1) the NASA Ocean Biology Processing Group Distributed Active Archive Center (OB.DAAC), which is responsible for data produced or collected under NASA's Earth Observing System Data and Information System but also

hosts ESA's S-3 data, and (2) the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA), which hosts land surface reflectance, surface temperature, and now provisional aquatic reflectance. Therefore, it is not generally feasible to attribute a portion of the initial setup costs to a single satellite product.

#### Sociodemographic distribution

Metropolitan and rural spatial information was obtained from the 2013 Rural-Urban Continuum Codes (USDA 2013). Federal and Tribal lands were identified in the Protected Areas Database of the USA (USGS 2016). We first look at the percentage of lakes with monitoring data within each of the datasets that fall within counties designated as metropolitan or rural areas according to the Bureau of Labor Statistics designations. Metropolitan areas have at least one urban core of 10,000 people or more in population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties.

To further characterize sociodemographic variation across the different datasets, we use county level data from the USDA Economic Research Service. These data on population, unemployment, poverty, and

income are derived from the US Census Bureau and Bureau of Labor Statistics data that are adjusted to provide consistent and reliable single-year estimates in both urban and rural areas (USDA 2017).

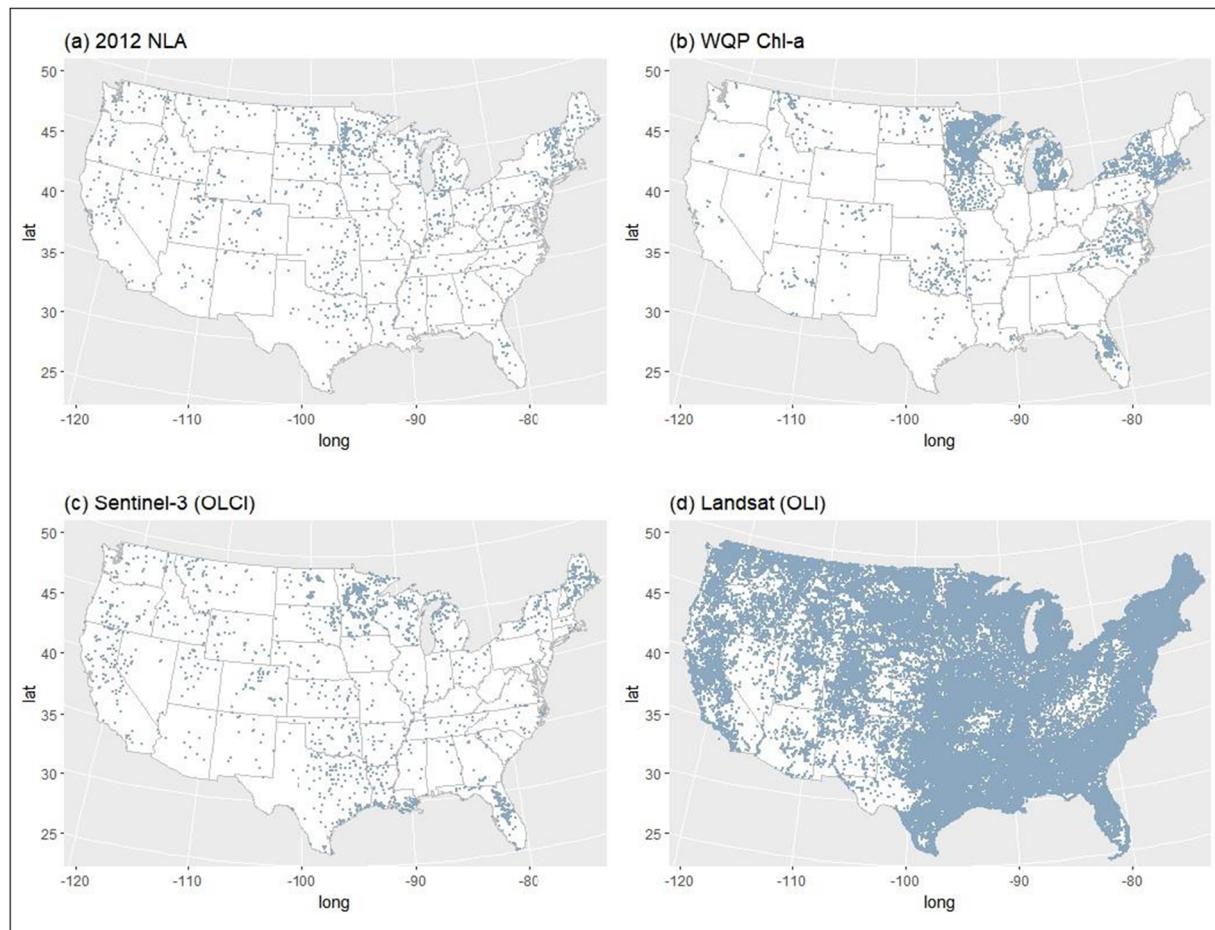
## Results and discussion

### Spatial and temporal coverage

The WQP returned a total of 36,165 in situ measures of chl-a from 1980 to 2015 (3,042 NWIS, 33,123 STORET). After selecting sites in the lower 48 states for which there were unique surface water chl-a sample locations (excluding the Great Lakes), there were 6265 unique waterbodies across the country. The 2012 NLA had 1269 sites with in situ measures of chl-a. Figure 3

shows the spatial coverage for each of these datasets. Each point on the map represents the waterbody centroid containing at least one in situ chl-a observation. As seen in the maps of satellite spatial coverage, L8 had the most coverage with over 170,000 waterbodies detectable relative to the 1857 waterbodies resolvable by S3. This difference is due to the 30-m pixel resolution available with L8 allowing it to resolve much smaller lakes relative to the subset of lakes resolvable with the 300-m pixel resolution for S3 (Clark et al. 2017).

The waterbodies included in the NLA data were selected to be sampled based on the probabilistic sampling design of the National Aquatic Resource Surveys (NARS) program under which the NLA is implemented. The advantage of the NLA data is that the NARS program's probabilistic design allows one to make assessments that are valid for reporting on nationwide and



**Fig. 3** Comparison of the number of lakes monitored in (a) 2012 National Lakes Assessment, (b) lakes that have been monitored for chl-a in situ with reporting in the National Water Quality Data

Portal, and those resolvable by satellite sensors: (c) Sentinel-3 (OLCI) and (d) Landsat-8 (OLI)

regional trends and is designed to be representative of waterbodies across the USA. One disadvantage of the NLA data is that data are only collected once every 5 years which limits the types of analysis and conclusions that it can support. How representative waterbodies detected by remote sensing are compared with the population of waterbodies across the nation is an open question for future research. The set of waterbodies that can be resolved by satellite sensors is biased based on waterbody size and shape, but this is dependent on the individual satellite sensor resolution. Table 1 provides summary statistics for the sampled lakes included in each dataset.

In addition to examining the data on a national scale, we also report the number of waterbodies resolved by each satellite platform in each state. Minnesota had the largest number of lakes (Fig. 4). In a little over half of the states, the number of waterbodies that had in situ data was about equal when comparing the S3 resolvable waterbodies and the 2012 NLA waterbodies. In states with larger numbers of waterbodies, the gap between S3 and NLA increased. The L8 satellite resolved the largest number of observable waterbodies (Fig. 3 and Table 1). However, this does not mean that L8 provides more observations than S3 over time. The S3 satellite has a shorter revisit time than L8 and an order of magnitude more annual observations per lake was available from the S3 OLCI satellite sensor (see Fig. 5). This high temporal frequency of observations is an advantage for monitoring programs that have a need for monitoring related to either posting or lifting HABs advisories. However, L8 does resolve smaller waterbodies which may be an advantage for numerous monitoring

applications relevant to eutrophication, HABs, and nuisance algae. One could derive a longer historical record back to 1980 if multiple Landsat missions were combined in sequence. It is important to note that if multiple Landsat missions were used for historical observations, validation across the different sensors would be required, but that exercise is beyond the scope of this study.

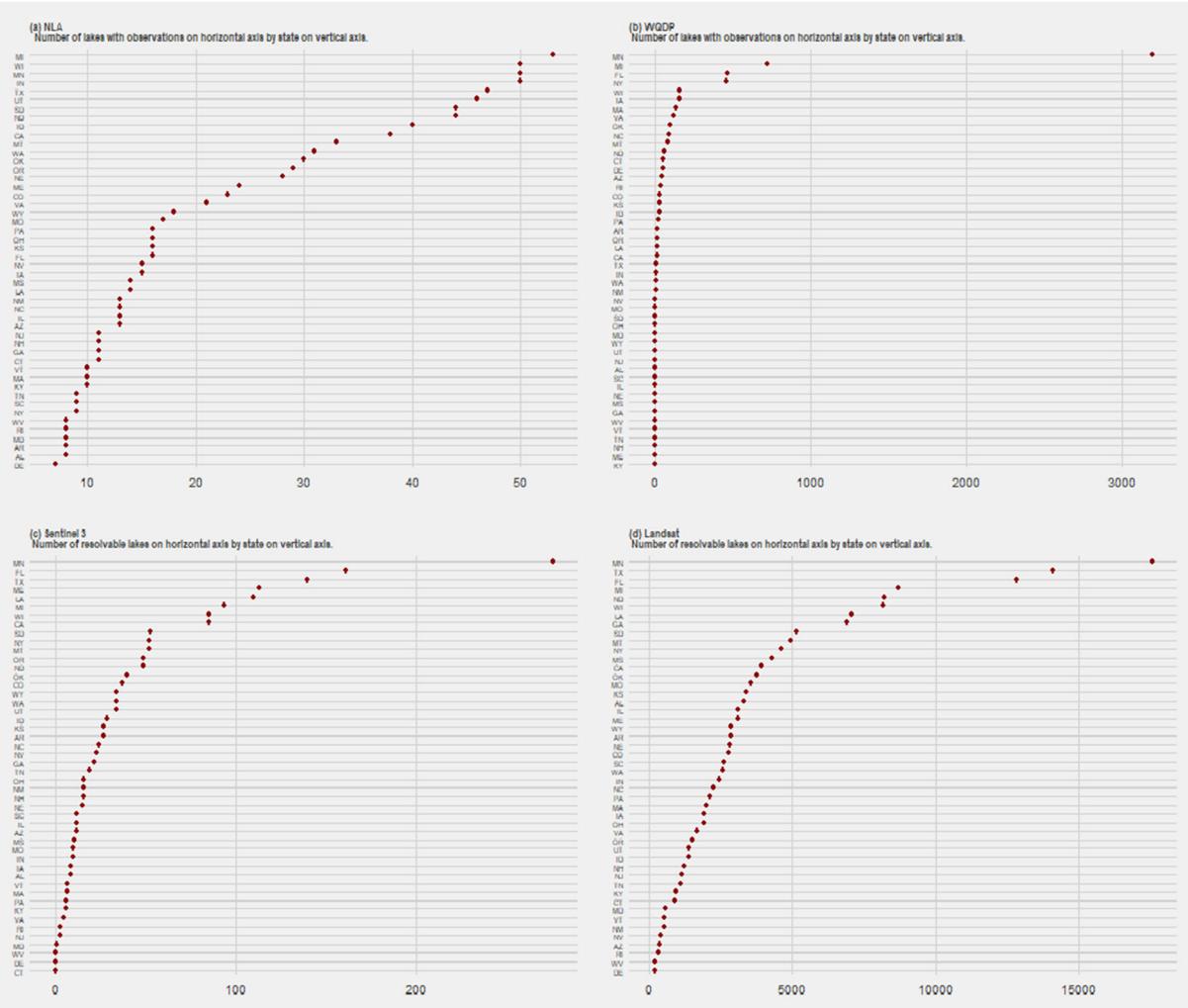
Figure 5 shows the distribution of annual observations per waterbody available from each satellite platform including cloud cover (Mercury et al. 2012) to provide the number of available annual observations. The number of observations per lake resolvable by L8 has a bimodal distribution of annual observations due to the widths of the satellite swaths. Where sides of swaths overlapped, the waterbodies within those overlapping swaths have a higher frequency of observations. While the L8 sensor does provide observations for more waterbodies and thus more total annual observations nationwide, S3 provided more annual observations per lake, with well over 150 observations available for each individual lake through the course of a year. The number of annual observations by S3 will increase with the combined monitoring of S3A and S3B recently launched in 2018. The high temporal frequency of these data is well-suited to monitoring programs focused on episodic events such as nuisance algae or HABs. For many states, S3 alone will provide large amounts of useful data to monitoring programs with > 1000 annual observations available in over 10 states. In terms of monitoring and research efforts focused on examining the causes or consequences of HAB events, the panel data structure available from satellite data (i.e., repeated measures on the same units) will allow researchers to exploit statistical methods that allow one to control for unobservable but fixed (time-invariant) confounding factors through panel designs. This is a significant advantage of this type of data relative to infrequently collected cross-sections of data (Hsiao 2007; Donaldson and Storeygard 2016).

Figure 6a shows a time series plot of total annual chl-a observations within the WQP between 1980 and 2015. There is a clear pattern with a steady increase in monitoring from 1980 to 2008. After 2008, there is a 50% drop in the number of chl-a observations relative to the number of chl-a observations available at the end of the evaluation period in

**Table 1** Summary of waterbodies included in each dataset

Lake observations	N	Total area	Median area	Max area	Min area
2012 NLA	1269	9814	0.273	1675	0.010
Sentinel-3 (OLCI)	1857	68,215	10.65	4309	1.60
Landsat-8 (OLI)	170,235	101,823	0.059	57,517	0.174
WQP	6265	150,639	0.587	57,517	0.004

Note: Areal units are square kilometers. The lake area summary statistics reflect the size of the sampled waterbodies but should not be interpreted as the area represented by the water quality samples



**Fig. 4** Comparison of lakes monitored in (a) 2012 National Lakes Assessment, (b) lakes that have been monitored for chl-a with reporting in the Water Quality Portal, and those resolvable by (c) Sentinel-3 (OLCI) sensor, and (d) Landsat-8 (OLI) sensor by state.

Some states participating in the NLA (such as MN and WI) chose to leverage NLA for state-scale survey by intensifying their sampling effort

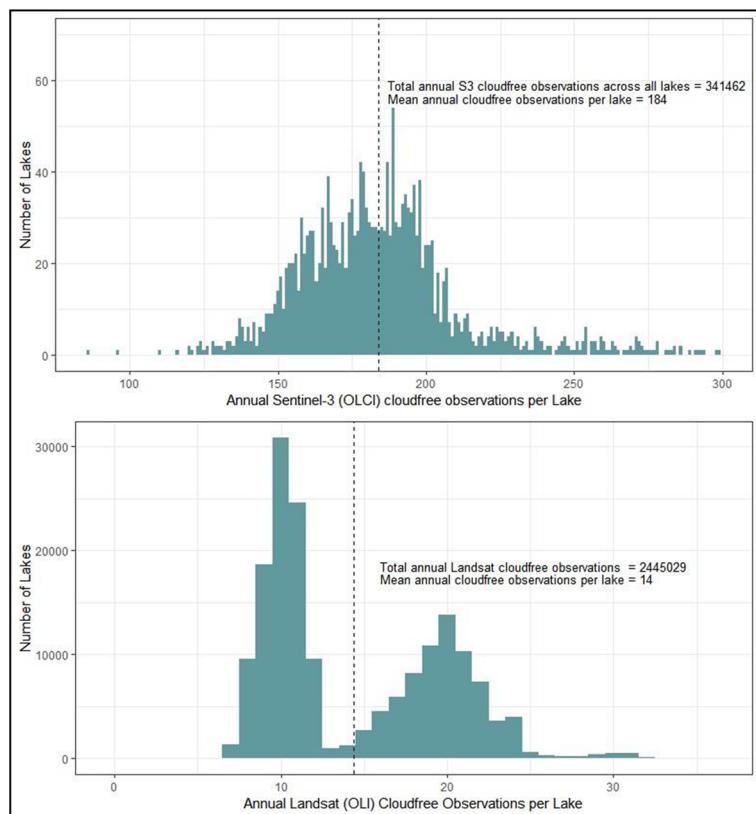
2015. The decline in observations may be due to a decline in voluntary reporting to online databases such as STORET, a decline in actual monitoring measurement efforts, or some combination of both. Seasonal sampling (Fig. 6b) is heavily biased toward warmer months with under-representation from November through March, typically the coldest months of the year. Figure 6c shows that most of the locations reported 1–25 observations between 2008 and 2015. In comparison with the number of observations available from the S3 satellite (over 341,000 annual observations across the nation), the potential for

satellite data to complement gaps in water quality monitoring efforts may prove valuable.

#### Satellite value-added information

The annual number of cloud-free observations for chl-a derived from S3 is approximately 341,000 cloud-free scenes (see Fig. 5). Using the assumed range of available scenes and cost parameters from Table 2, the distributions of estimated avoided costs in US dollars (\$USD) are shown in Fig. 7. These distributions reflect uncertainty in the parameters. The mean annual avoided

**Fig. 5** Distribution of annual cloud-free US lake and reservoir observations across Sentinel-3 (OLCI) and Landsat-8 (OLI) platforms



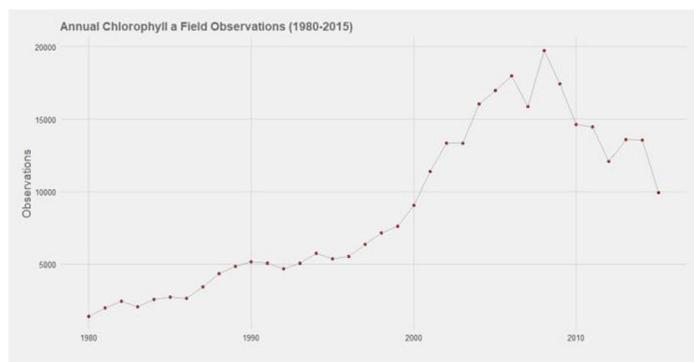
costs for the S3 are approximately \$5.7 million with a standard deviation of \$1.59 million. For L8, the mean annual avoided cost is approximately \$42 million with a standard deviation of \$9.5 million. The mean present value of a 10-year project using the satellite observations is approximately \$42 million ( $SD = \$9$  million) for S3 (OLCI) and \$316 million ( $SD = \$67$  million) for L8<sup>5</sup>. Because the mean annual expenditures required to make the data publicly available is largely driven by the cost of supplying two United States federal General Schedule (GS-13) scientists—the cost of storing and serving the data using cloud services is less than 1% of these labor costs—the scale of potential benefits of adopting satellite-derived measurements is substantial in comparison. These estimates are based on a hypothetical scenario of comparing a monitoring approach based solely on spatial and temporal scale of sampling that is available using the satellite data to one that is based on this same scale but requiring in situ sampling. This would

arguably never be the case, but these estimates do illustrate the large potential value is using the satellite data.

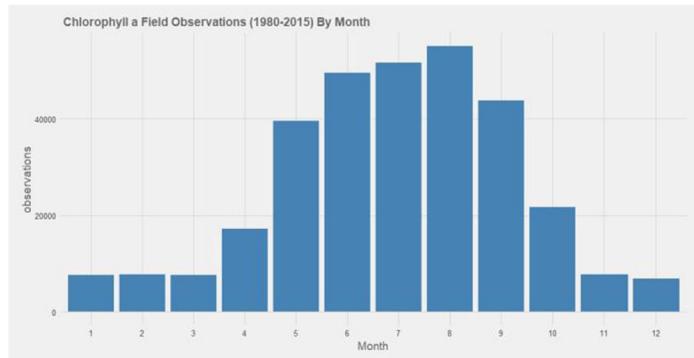
To highlight this point, we also consider a more realistic scenario under which a monitoring program chooses to rely mostly on in situ data collection with some augmentation with satellite data. For example, a state might choose to focus its in situ sampling program on waterbodies requiring less travel and time costs which constitute a large portion of the overall in situ sampling costs and only rely on the satellite data for more distant and remote waterbodies. To illustrate this idea, we compute the travel time and distance from each of the S3 waterbodies in California to the nearest of 13 California Water Resources Drinking Water district field offices. These are not necessarily the only field offices from which a monitoring program would send out field sampling teams, but we use these as an illustrative example. Figure 8 plots the empirical distribution function of travel times from each lake to the nearest field office. In an example scenario in which a monitoring program in California was to institute a consistent bi-weekly in situ sampling program of the 80% of the S3 lakes within a distance of roughly 150 mi from the

<sup>5</sup> The present value of these 10-year projects using a lower discount rate of 3% is \$50 million ( $SD = \$11$  million) for Sentinel-3 (OLCI) and \$370 million ( $SD = 79$  million) for Landsat (OLI).

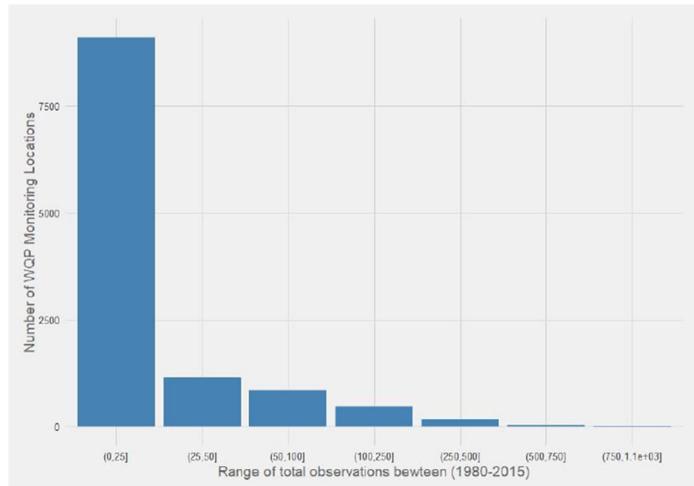
**Fig. 6** WQP observations as (a) the number of *in situ* observations by year, (b) total *in situ* samples by month, and (c) count of WQP monitoring locations binned by number of samples (2008–2015). Most sites have 1–25 samples in this timeframe



(a)



(b)



(c)

nearest field station and rely on the satellite information to monitor the remaining 20% of lakes, this would constitute a savings of about \$34,000 annually in terms of chl-a lab analysis costs at \$20 per analysis. If we additionally assume that the costs of sending out a field team to sample 4 lakes per day is \$200, then the annual savings would be approximately \$120,000. If one were

to extend this type of analysis and the underlying assumptions to the entire set of 1862 S3 resolvable lakes across the USA, the hypothetical cost savings would be approximately \$1,970,000 per year.

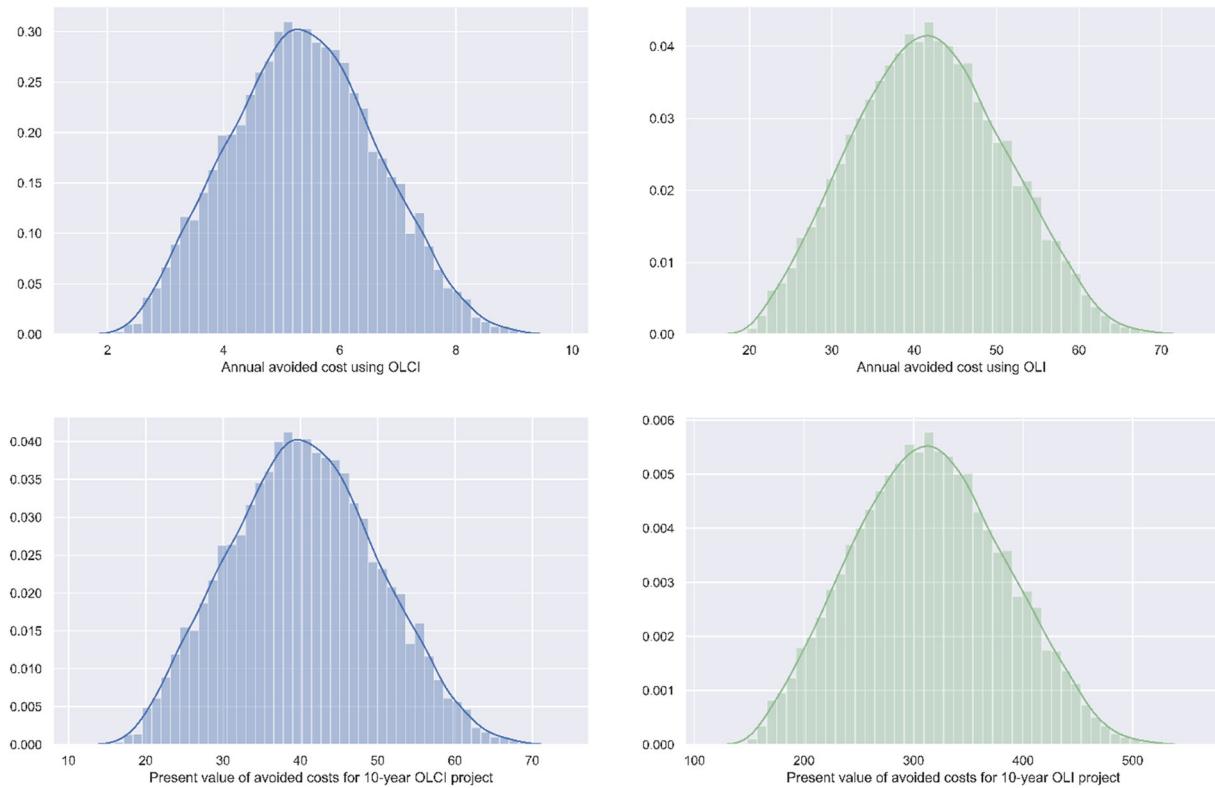
It is important to reiterate that using satellite data in no way obviates the need for *in situ* monitoring programs, as both systems provide relevant information at

**Table 2** Summary of cost variables used in avoided cost Monte-Carlo analysis

Variables included in avoided cost	Min	Mode	Max
Annual satellite observations (Sentinel-3 OLCI)	280,000	300,000	341,000
Annual satellite observations (Landsat-8 OLI)	1,833,000	1,956,000	2,445,000
Size of annual satellite data (GB)	4675	5500	6325
Annual cost per GB for cloud storage and bandwidth for input/output	\$3.8	\$4.8	\$5.75
Cost of single chl-a lab sample analysis	\$10	\$20	\$30
Annual labor cost for two GS-13 scientists	\$136,000	\$170,000	\$204,000

different spatial and temporal scales, specificity, accuracy, and precision. However, the above examples illustrate that even when monitoring programs complement a small portion of the in situ sampling potentially required for a continual HAB monitoring program, there are large potential cost saving associated with using the satellite data. That said, it is critical to keep in mind that satellites are limited in the variety of chemical, biological, and physical measures available, and systematic error in those measures can be a concern. In situ data

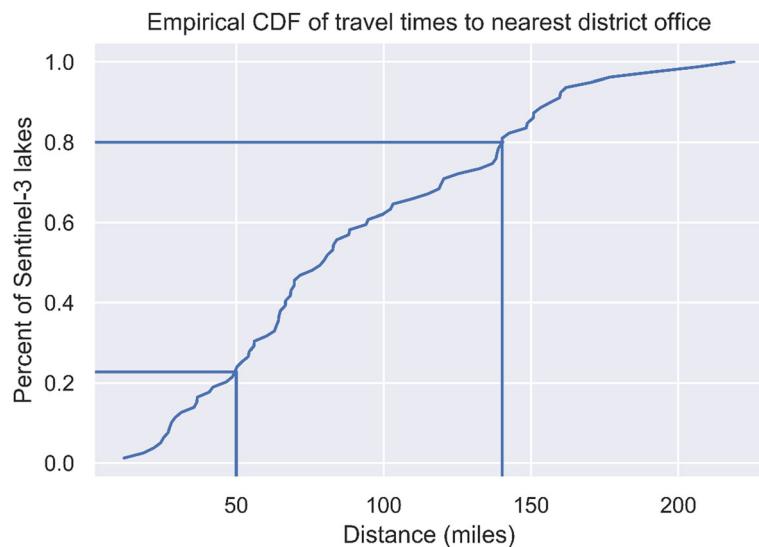
can be used to resolve these issues and are required for satellite algorithm validation (Greb et al. 2018). Although the S3 data do offer significant advantages in terms of temporal frequency, it is also important to remember that the S3 sensor does not have the spatial resolution to monitor smaller waterbodies, which constitute many lakes that are important in terms of local human activities and uses. Field monitoring also provides an opportunity to collect other environmental data that are not detectable by satellite yet can be important in



**Fig. 7** Distribution of potential avoided cost estimates in millions of \$USD. The first row is the annual estimate of avoided costs for the Sentinel-3 (OLCI) and Landsat-8 (OLI) platforms. The second

row is an estimate of the avoided costs estimated for a 10-year project discounted at 7%

**Fig. 8** Empirical cumulative distribution function (CDF) of travel distances from each of the Sentinel-3 waterbodies in California to nearest field office. On the y-axis, 0.2 can be interpreted as 20%



terms of a more complete understanding of complex aquatic processes. Both satellite and field sampling require analysis and interpretation. Despite these limitations to using the satellite data, the benefits of using this data over wide geographic regions in the USA and across the globe may be important. More advanced sensors and improved algorithms may serve to increase these advantages in the coming decade (Dekker et al. 2018; Greb et al. 2018).

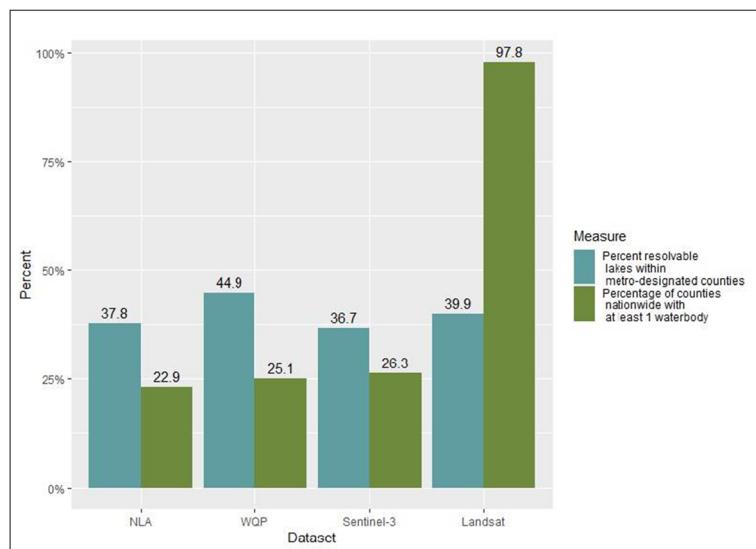
#### Sociodemographic attributes

In addition to the potential economic efficiencies associated with satellite-derived monitoring, it is also important to have some understanding of how the spatial distribution of water quality monitoring efforts intersects with socioeconomic characteristics of the communities surrounding those waterbodies. One of the primary purposes of water quality monitoring programs is to ascertain whether water quality is suitable for the intended use. For example, knowing how the temporal and spatial distribution of HAB events is distributed across jurisdictions and varying sociodemographic groups can help managers with the design of monitoring programs, identification of at-risk groups, and with the communication and translation of monitoring information to the public. In terms of HAB monitoring, identifying social and demographic data differences in these areas can be a useful step in developing management plans to reduce risk and exposure to nuisance and HAB events. It may also help to understand whether

monitoring programs are biased in ways that neglect specific communities. This can also help to identify different segments of the population that may benefit from satellite-derived water quality information and understand overall perceptions towards local environmental protection and management (McConnell 1997; Lo 2014). The cost efficiencies of augmenting monitoring programs with satellite-derived information is a consideration that water quality managers may want to consider in the context of their local needs. There are also additional benefits associated with improved monitoring in terms of providing information to the public that can help communities make better decisions regarding how they interact with potentially impaired waters, primarily in terms of avoiding exposure to toxic blooms (see Stroming et al. (2020) for a recent example of these types of avoidance behaviors).

Between 35 and 45% of the observations found in both the in situ and satellite datasets (Fig. 9) contain observations that are located within counties designated as metropolitan areas as defined by the United States Office of Management and Budget (USDA 2013). Just over 45% of monitored waterbodies in the WQP fall within counties designated as metropolitan. On one hand, enhanced monitoring within more populated areas could be important as these waterbodies may be more heavily used for different human activities and thus could be more of a risk for exposure to HAB events. However, in order to evaluate human exposure more accurately, one might utilize recreational time-use data to identify how much time humans spend at different

**Fig. 9** Percentage of lakes included in each dataset falling within counties designated as metropolitan areas (light blue) and the percentage of all counties nationwide with at least one observation

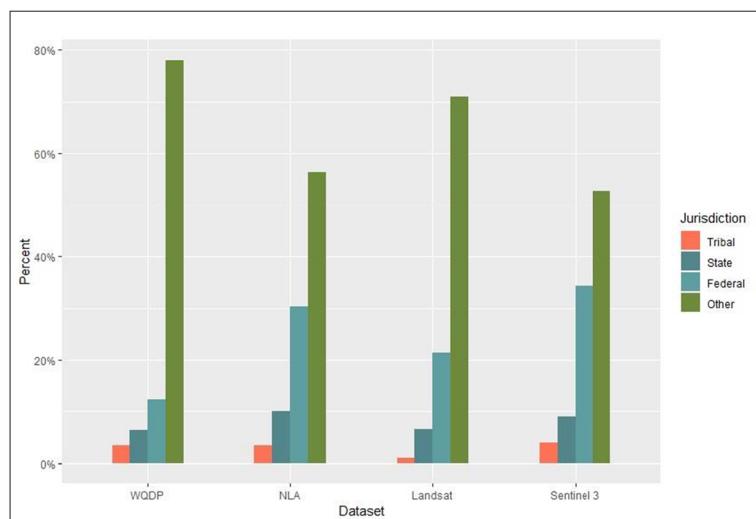


waterbodies. On the other hand, enhanced monitoring could be targeted toward landscapes with more nutrient run-off that might be more susceptible to HAB events. This is one simple example of the types of considerations that local managers might weigh as they determine which data resources are most relevant for their situation and needs.

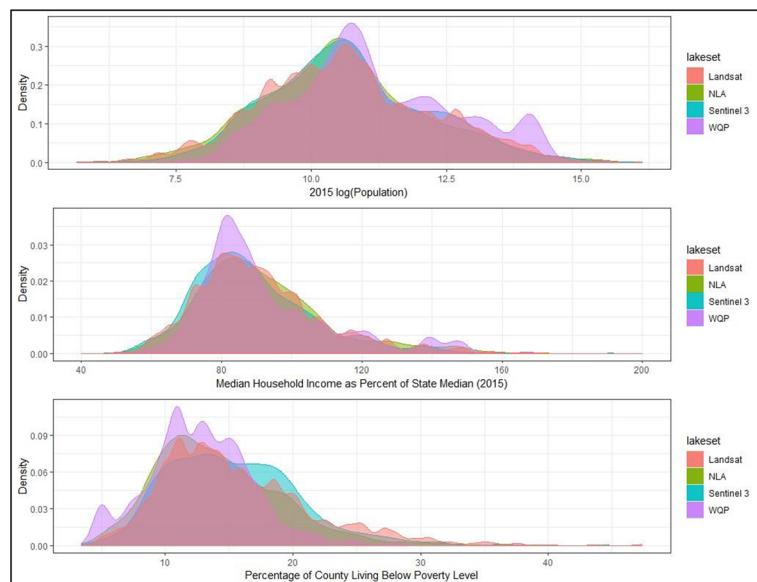
From a waterbody and watershed management perspective, it is also useful to know the land ownership patterns surrounding monitored waterbodies. Figure 10 charts the percentage of waterbodies from each dataset falling within federal, tribal, and “other” land ownership jurisdictions across the country. The “other” category includes waterbodies that were within a mixture of

public and private lands including state lands, county lands, and other private ownership patterns. The 2012 NLA monitoring data and S3 satellite data were nearly identical in terms of these percentages, while the L8 data had a smaller percentage of waterbodies on federal lands. This is largely because L8 resolves smaller waterbodies and has much denser spatial coverage than either the NLA or S3 data. The data from the WQP has the smallest percentage of waterbodies within federal lands, but this is consistent with the fact that data within the WQP was collected and submitted by state agencies and other non-federal organizations that monitor lakes within their management areas of interest. It is also consistent with the fact that more of the WQP

**Fig. 10** Percentage of lakes within each dataset across federal, state, tribal, and other land ownership jurisdictions



**Fig. 11** Comparing the distributions of socioeconomic outcomes in counties containing lakes across the different datasets. These kernel density plots are estimates of the probability density function and should be interpreted as smoothed histograms that are not as sensitive to the choice of bin size. The total area under the curve is 1 and the probability of any value lying between two points on the x-axis is given by the area under the curve between those points



waterbodies were located within metropolitan areas which are less likely to include federal lands. From a management perspective, both satellite datasets and the in situ datasets cover many areas that are not within tribal or federal jurisdiction. Having good monitoring data (including those from citizen science monitoring initiatives) in these areas can be beneficial for groups seeking to develop local management programs that address water issues which cut across state, county, and privately owned lands.

In Fig. 11, we examine the distribution of three socioeconomic variables across all counties containing monitoring data. Total population and median income were summarized as a percentage of the statewide median and the percentage of the county population living below the poverty line. For all these variables, there are no large visual discrepancies between the monitoring datasets in terms of how representative the locations are<sup>6</sup>. There is a slight skew toward larger populations in the WQP data, which is consistent with the slightly wider coverage of metropolitan counties as seen in Fig. 9. The WQP data also shows a slight skew to the left in the percentage of population living below the poverty line. The mean level is close to the 2015 national mean

of 13%, but there are fewer counties with poverty rates above the national mean relative to the other datasets.

In summary, when we examine a few select sociodemographic characteristics of the counties that encompass the water quality observations that are available across the satellite-based and in situ datasets, we find that most lakes are surrounded by private land with 30–40% of the observations falling within metropolitan-designated counties. Between 20 and 30% of counties across the lower 48 states have at least one observation from the different datasets apart from the L8 data which provides the most extensive coverage across the country due to its ability to resolve smaller waterbodies. There are few discrepancies across datasets in terms of county-level sociodemographic characterizing the counties with observations.

## Conclusions

Water quality monitoring is essential for maintaining aquatic resources that are safe and healthy for many uses such as drinking water, recreation, agriculture, and for maintaining the integrity of ecological resources that depend on healthy and well-functioning aquatic systems. Water quality monitoring provides critical information for understanding how well these goals are being met and to identify emerging concerns. Here, the focus was specifically on the availability of chl-a—one of the

<sup>6</sup> Two-sample nonparametric Kolmogorov-Smirnov statistics for each pairwise comparison of the socioeconomic variables across each dataset were performed to test whether each pair comes from the same distribution (Conover 1999). In all but two of the comparisons, the null hypothesis of equality in distributions is rejected. Results are available from the authors upon request.

primary indicators of phytoplankton biomass, eutrophication, nuisance algae, and HABs. We examined the potential availability and spatial coverage of observations across two satellite sensors L8 and S3, and across traditional field sampling datasets at the state-level (WQP) and the national level (2012 NLA). Neither of the traditional field monitoring datasets used in this comparison was intended or designed to be comprehensive HAB monitoring programs, but they represent the best data sources publicly available for the scale of our desired comparisons.

In addition to comparing the spatial and temporal coverage of these different data sources, we computed the potential value of utilizing satellite data (as an avoided cost) to complement existing water quality monitoring efforts. While we are aware of limited (Stroming et al. 2020) studies examining the value of satellite data for water quality monitoring, USGS did conduct a survey of Landsat imagery users to examine their uses of the data and to evaluate the mean economic value to those users employing a contingent valuation survey (USGS 2013). This study found that the total annual economic benefits from the Landsat imagery was over \$1.7 billion for US users alone, with a mean value of \$900 per imagery scene for a single user. Those reported values only include registered users of the data and do not include the additional benefits of products based on the imagery, such as using it to develop monitoring programs for chl-a as is examined in this study which range from \$5.7 to \$316 million depending on the satellite platform and timeframe. Our results complement these numbers in that they indicate the high value of information provided by satellites and remote sensing technology. As satellite data is incorporated into more water quality monitoring and management programs, there will be more opportunities to further examine and quantify the value of this information by collecting data on how the information is being used to change management decisions, monitoring program activities, and how the public responds to this information in comparison to choices made in the absence of this information.

Satellite-derived chl-a data complement traditional in situ monitoring in terms of filling temporal and spatial gaps. It is important to reiterate that in situ

monitoring programs provide important data that are necessary to validate water quality measures based on remote sensed imagery. In situ data are also currently the only method for determining toxin concentrations and provide other co-located data (e.g., nutrient concentrations) that can be used to refine ecological understanding and processes within systems that inform management (Yuan et al. 2014; Yuan and Pollard 2015). While the L8 data provide the best geographic coverage for most US lakes and reservoirs, the S3 data provide temporal coverage not available from any other data source. Access to these synoptic spatial and temporal data is particularly useful for monitoring and studying water quality problems such as nuisance algae and HABs that are more episodic in nature. Even without this high temporal frequency, L8 data provide an opportunity to access information difficult to obtain by traditional field sampling due to the expansive geographic coverage and extended time-series dating back to 1980. In both cases of the satellite data, the availability of repeated measures across a large spatial scale is an enormous advantage for scientists and researchers. The analysis of repeated measures or panel data allows for a more robust set of empirical methods to control for biases and unmeasured confounding that plague many studies using only a single cross-section of data. The improved monitoring capacity may also be utilized by states to monitor water quality impairments and detection of pollution events.

The automated nature of collecting satellite images, together with cloud-based computing platforms for processing, communicating, and translating this data into useful information, has enormous potential to improve water quality monitoring programs. Clearly, the cost of providing this satellite data to the public is relatively small compared with the cost that would be required to obtain the same information and data from traditional in situ methods. This is particularly relevant in areas that cannot implement or develop monitoring programs due to various constraints or during periods, such as the Coronavirus pandemic, or other unforeseen events that may inhibit traditional in situ data collection. Additionally, this study demonstrated that the satellite data provide good coverage in terms of monitoring waterbodies that are not overly biased toward a particular set of communities differentiated

by the small set of socioeconomic characteristics examined here. Future work could examine in more detail how the economic benefits associated with improved monitoring may vary across communities. For example, rural communities may benefit relatively more than urban communities since remote sensing data might be the *only* information they have if their community lacks a formal water in situ monitoring program. On the other hand, more urban, densely populated areas may require more frequent, spatially comprehensive data because there are likely to be more intensive interactions between humans and lakes in these areas. These are important and relevant questions that will require specific data on how humans respond to and use monitoring information. Only by pursuing these types of studies will we be able to comprehensively measure and quantify the value of information derived from remote sensing programs. These values which remain largely unmeasured, in addition to the values computed in this study, help to shape priorities and investments in remote sensing technology and begin to illuminate the vast economic value that remote sensing technology provides to society. While there may be clear financial incentives to adopt usage of satellite-derived data, one must also remember that these are not the only benefits. As algorithms to interpret satellite data improve, the representativeness, accuracy, and faster data generation periods should improve scientific understanding and awareness relevant to protecting the environment.

This work was supported by the National Aeronautics and Space Administration (NASA) Ocean Biology and Biogeochemistry Program/Applied Sciences Program (proposal 14-SMDUNSOL14-0001), the United States Environmental Protection Agency (USEPA), the National Oceanic and Atmospheric Administration (NOAA), and the United States Geological Survey (USGS) Toxic Substances Hydrology Program. This article has been reviewed by the National Health and Environmental Effects Research Laboratory and approved for publication. We thank several anonymous reviewers for their contributions. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the US Government. The views expressed in this article are those solely of the authors from USEPA and do not necessarily reflect the views or policies of the USEPA but do represent the views of the USGS.

## References

- Anderson, D. M. (2009). Approaches to monitoring, control and management of harmful algal blooms (HABs). *Ocean & Coastal Management, Safer Coasts, Living with Risks: Selected Papers from the East Asian Seas Congress 2006, Haikou, Hainan, China*, 52(7), 342–347. <https://doi.org/10.1016/j.ocecoaman.2009.04.006>.
- Anderson-Abbs, B., Howard, M., Taberski, K., and Worcester, K. (2016). California freshwater harmful algal blooms assessment and support strategy. California Water Boards; [https://www.waterboards.ca.gov/sanfranciscobay/water\\_issues/programs/SWAMP/HABstrategy\\_phase%201.pdf](https://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/SWAMP/HABstrategy_phase%201.pdf). Accessed 10 Oct 2020.
- Backer, L., Landsberg, J. H., Miller, M., Keel, K., & Taylor, T. (2013). Canine cyanotoxin poisonings in the United States (1920s–2012): Review of suspected and confirmed cases from three data sources. *Toxins*, 5(9) Multidisciplinary Digital Publishing Institute, 1597–1628.
- Brooks, B. W., Lazorchak, J. M., Howard, M., Johnson, M. V., Morton, S. L., Perkins, D., Reavie, E. D., Scott, G. I., Smith, S. A., & Steevens, J. A. (2016). Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environmental Toxicology and Chemistry*, 35(1) Wiley Online Library, 6–13.
- Chipman, J. W., Lillesand, T. M., Schmaltz, J. E., Leale, J. E., & Nordheim, M. J. (2004). Mapping lake water clarity with Landsat images in Wisconsin, U.S.A. *Canadian Journal of Remote Sensing*, 30(1) Taylor & Francis, 1–7. <https://doi.org/10.5589/m03-047>.
- City of Salem, 2018. Drinking Water Advisory. Accessed August 10, 2020. <https://www.cityofsalem.net/Pages/drinking-water-advisory-may-29.aspx>.
- Clark, J. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., Johnston, J. M., Ignatius, A. R., Myer, M. H., Loftin, K. A., Werdell, P. J., & Stumpf, R. P. (2017). Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. *Ecological Indicators*, 80. Elsevier, 84–95.
- Codd, G. A. (2000). Cyanobacterial toxins, the perception of water quality, and the prioritisation of eutrophication control. *Ecological Engineering*, 16(1), 51–60. [https://doi.org/10.1016/S0925-8574\(00\)00089-6](https://doi.org/10.1016/S0925-8574(00)00089-6).
- Conover, W. (1999). *Practical Nonparametric Statistics* (3rd ed.). New York: Wiley.
- Dekker, A.G., Pinnel, N., Gege, P., Briottet, X., Court, A., and Steef, P. (2018). Feasibility study for an aquatic ecosystem Earth observing system. <https://www.geoaquawatch.org/wp-content/uploads/2017/08/CEOS-Report-vs2-5-April-2018-low-res.pdf>. Accessed 10 Oct 2020.
- Devlin, M., Bricker, S., & Painting, S. (2011). Comparison of five methods for assessing impacts of nutrient enrichment using estuarine case studies. *Biogeochemistry*, 106(2) Springer, 177–205.
- Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *The Journal of Economic Perspectives*, 30(4), 171–198.
- Dörnhöfer, K., & Oppelt, N. (2016). Remote sensing for lake research and monitoring—recent advances. *Ecological Indicators*, 64. Elsevier, 105–122.

- ESOV. (2017). Earth observation swath and orbit visualisation tool. <http://eop-cfi.esa.int/index.php/applications/esov>. Accessed 10 Oct 2020.
- Ferreira, J. G., Andersen, J. H., Borja, A., Bricker, S. B., Camp, J., Da Silva, M. C., Garcés, E., Heiskanen, A. S., Humborg, C., Ignatiades, L., & Lancelot, C. (2011). Overview of eutrophication indicators to assess environmental status within the European marine strategy framework directive. *Estuarine, Coastal and Shelf Science*, 93(2) Elsevier, 117–131.
- Franz, B. A., Bailey, S. W., Kuring, N., & Werdell, P. J. (2015). Ocean color measurements with the Operational Land Imager on Landsat-8: implementation and evaluation in SeaDAS. *Journal of Applied Remote Sensing*, 9(1), 096070.
- Getelson, A. (1992). The peak near 700 NM on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing*, 13(17) Taylor & Francis, 3367–3373.
- Gomarasca, M. A., Giardino, C., Bresciani, M., De Carolis, G., Sandu, C., Tornato, A., Spizzachino, D., Valentini, E., TaraMelli, A., & Tonolo, F. (2019). Copernicus Sentinel missions for water resources. In *Proceedings of 6th International Conference on Space Science and Communication*. Johor: Malaysia.
- Graham, J. L., Loftin, K. A., & Kamman, N. (2009). Monitoring recreational freshwaters. *Lakelines*, 29, 18–24.
- Greb, S., Odermatt, D., Schaeffer, B. A., Spyrikos, E., and Wang, M. (2018). Complementarity of in situ and satellite measurements. In *Earth Observations in Support of Global Water Quality Monitoring*, edited by S. Greb, A. Dekker, and C. Binding, 29–42. IOCCG Report Series 17. Dartmouth, Canada: International Ocean Colour Coordinating Group.
- Heisler, J., Gilber, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., Dortch, Q., Gobler, C. J., Heil, C. A., Humphries, E., & Lewitus, A. (2008). Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae, HABs and Eutrophication*, 8(1), 3–13. <https://doi.org/10.1016/j.hal.2008.08.006>.
- Ho, J. C., & Michalak, A. M. (2015). Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie. *Journal of Great Lakes Research*, 41(2) Elsevier, 317–325.
- Hsiao, C. (2007). Panel data analysis – advantages and challenges. *Test*, 16(1), 1–22.
- Krumm, N., and N. Hoffman. (2020). Practical estimation of cloud storage costs for clinical genomic data. *Practical Laboratory Medicine* 21. <https://doi.org/10.1016/j.plabm.2020.e00168>
- Kutser, T., Metsamaa, L., Strömbeck, N., & Vahtmäe, E. (2006). Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science*, 67(1), 303–312. <https://doi.org/10.1016/j.ecss.2005.11.024>.
- Lee, A. (2014). Monte Carlo Error Propagation. Retrieved from: <https://pythonhosted.org/mcerp/index.html>. Accessed 10 Oct 2020.
- Lo, A. Y. (2014). Negative income effect on perception of long-term environmental risk. *Ecological Economics*, 107. Elsevier, 51–58.
- Loftin, K. A., Graham, J. L., Hilborn, E. D., Lehmann, S. C., Meyer, M. T., Dietze, J. E., & Griffith, C. B. (2016). Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. *Harmful Algae*, 56(June), 77–90. <https://doi.org/10.1016/j.hal.2016.04.001>.
- McConnell, K. E. (1997). Income and the demand for environmental quality. *Environment and Development Economics*, 2(4) Cambridge University Press, 383–399.
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Reah, A. (2012). NHDPlus Version 2: User Guide. [https://nhdplus.com/NHDPlus/NHDPlusV2\\_documentation.php](https://nhdplus.com/NHDPlus/NHDPlusV2_documentation.php). Accessed 10 Nov 2020.
- Mercury, M., Green, R., Hook, S., Oaida, B., Wu, W., Gudnerson, A., & Chodas, M. (2012). Global cloud cover for assessment of optical satellite observation opportunities: A Hyperspectral case study. *Remote Sensing of Environment*, 126. Elsevier, 62–71.
- Morgan, M. , Henrion, M., and Small, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press.
- Mouw, C. B., Greb, S., Aurin, D., DiGiacomo, P. M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., & Moses, W. (2015). Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sensing of Environment*, 160(April), 15–30. <https://doi.org/10.1016/j.rse.2015.02.001>.
- North American Lake Management Society (NALMS) (n.d.). Chlorophyll analysis. Retrieved from <https://www.nalms.org/secchidipin/monitoring-methods/chlorophyll-analysis/>. Accessed 10 Oct 2020.
- Pahlevan, N., Schott, J. R., Franz, B. A., Zibordi, G., Markham, B., Bailey, S., Schaaf, C. B., Ondrusek, M., Greb, S., & Strait, C. M. (2017). Landsat 8 remote sensing reflectance ( $R_{rs}$ ) products: Evaluations, intercomparisons, and enhancements. *Remote Sensing of Environment*, 190, 289–301.
- Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Ha, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M., Ondrusek, M., Oppelt, N., & Stumpf, R. (2020). Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and S3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environment*, 240(February), 111604. <https://doi.org/10.1016/j.rse.2019.111604>.
- Palmer, S. C. J., Kutser, T., & Hunter, P. D. (2015). Remote sensing of inland waters: challenges, progress and future directions. *Remote Sensing of Environment*, 157(February), 1–8. <https://doi.org/10.1016/j.rse.2014.09.021>.
- Pollard, A. I., Hampton, S. E., & Leech, D. M. (2018). Using the U.S. Environmental Protection Agency's National Lakes Assessment. *Limnology and Oceanography Bulletin*, 27(2), 36–41.
- Rajakumari, S., Nalini, B., & Nalini, C. (2014). An efficient cost model for data storage with horizontal layout in the cloud. *Indian Journal of Science and Technology*, 7(March), 45–46. <https://doi.org/10.17485/ijst/2014/v7i3S/48655>.
- Read, E. K., Carr, L., De Cicco, L., Dugan, H. A., Hanson, P. C., Hart, J. A., Kreft, J., Read, J. S., & Winslow, L. A. (2017). Water quality data for national-scale aquatic research: The Water Quality Portal. *Water Resources Research*, 53(2) Wiley Online Library, 1735–1745.
- Schaeffer, B. A., Iiames, J., Dwyer, J., Urquhart, E., Salls, W., Rover, R., & Seegers, B. (2018a). An initial validation of Landsat 5 and 7 derived surface water temperature for US

- lakes, reservoirs, and estuaries. *International Journal of Remote Sensing*, 39(22), 7789–7805.
- Schaeffer, B. A., Schaeffer, K. G., Keith, D., Lunetta, R. S., Conmy, R., & Gould, R. W. (2013). Barriers to adopting satellite remote sensing for water quality management. *International Journal of Remote Sensing*, 34(November), 7534–7544. <https://doi.org/10.1080/01431161.2013.823524>.
- Schaeffer, B. A., Bailey, S. W., Conmy, R. N., Galvin, M., Ignatius, A. R., Johnston, J. M., Keith, D. J., Lunetta, R. S., Parmar, R., Stumpf, R. P., & Urquhart, E. A. (2018b). Mobile device application for monitoring cyanobacteria harmful algal blooms using S3 satellite ocean and land colour instruments. *Environmental Modelling & Software*, 109. Elsevier, 93–103.
- Schaeffer, B., Hagy III, J., Conmy, R., Lehrter, J., & Stumpf, R. (2011). An approach to developing numeric water quality criteria for coastal waters using the SeaWiFS satellite data record. *Environmental Science & Technology*, 46(December), 916–922. <https://doi.org/10.1021/es2014105>.
- Schaeffer, B., Loftin, K., Stumpf, R., & Werdell, J. (2015). Agencies collaborate, develop a cyanobacteria assessment network. *Eos*, 96(November). <https://doi.org/10.1029/2015EO038809>.
- Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., & Werdell, P. J. (2018). Performance metrics for the assessment of satellite data products: an ocean color case study. *Optics Express*, 26(2), 7404–7422.
- Smayda, T. (1997). Harmful algal blooms: their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology and Oceanography*, 42(5part2), 1137–1153.
- Smith, V. H., & Schindler, D. W. (2009). Eutrophication science: Where do we go from here? *Trends in Ecology & Evolution*, 24(4), 201–207. <https://doi.org/10.1016/j.tree.2008.11.009>.
- Smith, V., Dodds, W., Havens, K., Engstrom, D. R., Paerl, H., Moss, B., & Likens, G. E. (2014). Comment: Cultural eutrophication of natural lakes in the United States is real and widespread. *Limnology and Oceanography*, 59(November), 2217–2225. <https://doi.org/10.4319/lo.2014.59.6.2217>.
- State of Ohio. (2016) Harmful algal bloom response strategy for recreational waters. Retrieved from <https://www.epa.state.oh.us/portals/35/hab/HABResponseStrategy.pdf>. Accessed 10 Oct 2020.
- Stewart, I., Webb, P. M., Schluter, P. J., & Shaw, G. R. (2006). Recreational and occupational field exposure to freshwater cyanobacteria – a review of anecdotal and case reports, epidemiological studies and the challenges for epidemiologic assessment. *Environmental Health*, 5(1), 6. <https://doi.org/10.1186/1476-069X-5-6>.
- Stroming, S., Robertson, M., Mabee, B., Kuwayama, Y., and 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*, e2020GH000254.
- Stumpf, R. P., Davis, T. W., Wynne, T. T., Graham, J. L., Loftin, K. A., Johengen, T. H., Gossiaux, D., Palladino, D., & Burtner, A. (2016). Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae*, *Global Expansion of Harmful Cyanobacterial Blooms: Diversity, ecology, causes, and controls*, 54(April), 160–173. <https://doi.org/10.1016/j.hal.2016.01.005>.
- Stumpf, R., Culver, M. E., Tester, P., Tomlinson, M., Kirkpatrick, G., Pederson, B. A., Truby, E., Ransibrahmanakul, V., & Soracco, M. (2003). Monitoring Karenia Brevis blooms in the Gulf of Mexico using satellite ocean color imagery and other data. *Harmful Algae*, 2(June), 147–160. [https://doi.org/10.1016/S1568-9883\(02\)00083-5](https://doi.org/10.1016/S1568-9883(02)00083-5).
- Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sensing*, 8(8), 640. <https://doi.org/10.3390/rs8080640>.
- Trees, C., Kennicutt, M., & Brooks, J. (1985). Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments. *Marine Chemistry*, 17(1), 1–12. [https://doi.org/10.1016/0304-4203\(85\)90032-5](https://doi.org/10.1016/0304-4203(85)90032-5).
- Tyler, A. N., Hunter, P. D., Spyros, E., Groom, S., Constantinescu, A. M., & Kitchen, J. (2016). Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal, and shelf-sea waters. *Science of the Total Environment*, 572(December), 1307–1321. <https://doi.org/10.1016/j.scitotenv.2016.01.020>.
- USDA. (2013). Rural-urban continuum codes. <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/>. Accessed 10 Oct 2020.
- USDA. (2017). Economic research service county-level data sets. <https://www.ers.usda.gov/data-products/county-level-data-sets/documentation/>. Accessed 10 Oct 2020.
- USDA-STEWARDS. (2017). USDA sustaining the Earth's watersheds - agricultural research data system STEWARDS. Retrieved from <https://www.nrrig.mwa.ars.usda.gov/stewards/stewards.html>. Accessed 10 Oct 2020.
- USEPA. (2014). Guidelines for preparing economic analyses. Retrieved from <https://www.epa.gov/sites/production/files/2017-08/documents/ee-0568-50.pdf>. Accessed 10 Oct 2020.
- USEPA. (2016). National lakes assessment 2012: A collaborative survey of lakes in the United States. Retrieved from <https://nationallakesassessment.epa.gov/>. Accessed 10 Oct 2020.
- USEPA. (2017). Storage and retrieval data System. Retrieved from: <https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange>. Accessed 10 Oct 2020.
- USGS. (2013). Users, uses, and value of landsat satellite imagery - Results from the 2012 survey of users. <https://eros.usgs.gov/doi-remote-sensing-activities/2016/users-uses-and-value-landsat-imagery>. Accessed 10 Oct 2020.
- USGS. (2016). Protected areas database of the United States (PAD-US), version 1.4 combined feature class. <https://gapanalysis.usgs.gov/padus/data/>. Accessed 10 Oct 2020.
- USGS. (2017a). USGS water data for the Nation. *National Water Information System database*. <https://doi.org/10.5066/F7P55KJN>.
- USGS. (2017b). USGS land acquisition tool. [https://landsat.usgs.gov/landsat\\_acq](https://landsat.usgs.gov/landsat_acq). Accessed 10 Oct 2020.
- USGS, USEPA, and National Water Quality Monitoring Council. (2017). Water quality portal. <https://www.waterqualitydata.us/>. Accessed 10 Oct 2020.
- Wainwright, J., & Mulligan, M. (2013). *Environmental modelling: Finding simplicity in complexity* (p. 496). UK: John Wiley and Sons.

- 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. *Limnology and Oceanography*, 55(5) Wiley Online Library, 2025–2036.
- 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(12) Taylor & Francis, 3665–3672.
- Yuan, D., Cui, L., Liu, X., Fu, E., and Yang, Y. (2016). A cost-effective strategy for storing scientific datasets with multiple service providers in the cloud. *ArXiv E-Prints*, January.
- Yuan, L., and Pollard, A. I. (2015). Deriving nutrient targets to prevent excessive cyanobacterial densities in U.S. lakes and reservoirs. *Freshwater Biology* 60 (June). doi:<https://doi.org/10.1111/fwb.12620>.
- Yuan, L., Pollard, A. I., Pather, S., Oliver, J. L., & D'Anglada, L. (2014). Managing microcystin: Identifying national-scale thresholds for total nitrogen and chlorophyll-a. *Freshwater Biology*, 59(June), 1970–1981. <https://doi.org/10.1111/fwb.12400>.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.