Best Practices to Employ Satellite Remote Sensing to Assess Freshwater Harmful Algal Blooms in Inland Waters of California

Jayme Smith
Marisa Van Dyke
Carly Nilson
Alle Lie
Martha Sutula

SOUTHERN CALIFORNIA COASTAL WATER RESEARCH PROJECT

Best Practices to Employ Satellite Remote Sensing to Assess Freshwater Harmful Algal Blooms in Inland Waters of California

Jayme Smith[1], Marisa Van Dyke[2], Carly Nilson[2], Alle Lie[1], and Martha Sutula[1]

[1] Southern California Coastal Water Research Project, Costa Mesa, CA

[2] State Water Resources Control Board, Sacramento, CA

March 2025

Acknowledgements

The authors wish to thank the members of the Technical Advisory Committees (TAC) and other participants for assisting with the development of content for this document and reviewing draft versions. The Committee members are listed below in alphabetical order. Additional reviewers not on the TAC provided extensive comments and advice, for which we are profoundly grateful. TG = Technical Working Group, UG = User Working Group.

Name	Institute	TG	UG
Keith Bouma-Gregson	United States Geological Survey	X	
Stefan Cajina	State Water Resources Control Board		X
Megan Coffer	National Oceanic and Atmospheric Administration	X	
Ali Dunn	State Water Resources Control Board		X
Daniel Ellis	State Water Resources Control Board		X
Katie Fong	State Water Resources Control Board		X
Meredith Howard	Central Valley Regional Water Quality Control Board		X
Carey Kowalski	San Diego Regional Water Quality Control Board		X
Jeanie Mascia	State Water Resources Control Board		X
Diana Messina	State Water Resources Control Board		X
Karen Odkins	California Department of Fish and Wildlife		X
Bridget Seegers	MSU/NASA Goddard Space Flight Center	X	
Michelle Tomlinson	National Oceanic and Atmospheric Administration	X	
Randy Turner	San Francisco Estuary Institute		X

Prepared for:

As a Deliverable for the State Water Quality Control Board “Freshwater and Estuarine Harmful Algal Bloom Work Plan”

(Agreement No. 19-078-270-3)

This report should be cited as:

Smith, J., M. Van Dyke, C. Nilson, A. Lie, and M. Sutula. 2024. Assessment Approaches for Freshwater Harmful Algal Blooms in Inland Waters using Satellite Remote Sensing. Southern California Coastal Water Research Project. Costa Mesa, CA. www.sccwrp.org

Keywords:

Harmful algal blooms; cyanotoxins; cyanobacteria; recreation; lakes; reservoirs

List of Acronyms

AB 834	Assembly Bill 834 Freshwater and Estuarine Harmful Algal Bloom Program
CCHAB	California Cyanobacteria Harmful Algal Bloom Network
CAWB	California Water Boards
Chl-a	Chlorophyll a
CIcyano	Modified Cyanobacteria index
CyAN	Cyanobacteria Assessment Network
EPA	Environmental Protection Agency
DQI	Data quality indicator
ESA	European Space Agency
FHAB	Freshwater harmful algal bloom
GUI	Graphical user interface
MERIS	MEdium Resolution Imaging Spectrometer
MQO	Measurement quality objective
NASA	National Aeronautics and Space Administration
NHD	National Hydrography Dataset
NLCD	National Land Cover Database
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
OIMA	Office of Information Management and Analysis
OLCI	Ocean and Land Colour Instrument
QA/QC	Quality Assurance/Quality Control
R	R statistical program
RWQCB	Regional Water Quality Control Board
SCCWRP	Southern California Coastal Water Research Project Authority
SFEI	San Francisco Estuary Institute
SWAMP	Surface Water Ambient Monitoring Program
SWRCB	State Water Resources Control Board
TAC	Technical Advisory Committee

USACE	United States Army Corps of Engineer
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey

EXECUTIVE SUMMARY

Satellite imagery is a powerful tool that can enhance the monitoring and assessment of many of the lakes, reservoirs, and large estuaries in California. The synoptic view that satellite imagery provides makes it particularly useful for assessing freshwater harmful algal bloom events (FHABs), which can be either chronic or episodic in nature. FHABs negatively impact multiple beneficial uses in surface waters, including drinking water, recreation, tribal and cultural uses, fishing, agriculture, and aquatic life and can occur in all types of waterbodies. The degradation of these uses has broad and sustained ecosystem and economic impacts. However, little is understood about the extent of FHAB risks to beneficial uses because many of California's freshwaters are not routinely monitored for these impacts. The overall goal of this report is to support the broader applications of satellite imagery for assessments of FHABs. Wider use of this technology has been constrained due to the limited understanding of the quality of data and standardized procedures are lacking for processing and analyzing satellite imagery for additional analysis. Establishing these best practices and quantifying data quality opens the door to routinely use satellite imagery, as a complement to other data sources, to assess HABs in California's freshwater and estuarine waters as well as support water quality management decisions.

Towards this goal, this report encompasses the following topics with a focus on use of the Ocean and Land Color Instrument (OLCI) data products from onboard the Sentinel-3 satellite mission: 1) Identify management questions of relevance to remote sensing of FHABs, 2) Identify specific metrics that can be used to answer management questions, 3) Develop and document "best practices" protocols for processing raw remotely sensed imagery with supporting open-source code; 4) Develop specific procedures to calculate and visualize metrics used to answer management questions, with supporting open-source code; 5) Conduct a literature review to support Water Board decisions on quality assurance and quality control criteria; and 6) Identify key gaps in existing products and make strategic recommendations towards filling these gaps. Each of these elements were developed in collaboration with two advisory groups, one group with technical expertise in remote sensing and the other with expertise in water quality and FHAB assessment that were identified as likely end users of the technical report and related products.

Major Findings/Summary

A common set of water quality assessment metrics that can be addressed with OLCI satellite imagery data were identified by a broad array of stakeholders. On a broad level, these assessments are used to answer questions focused on status, trends, environmental drivers, and FHAB incident response. The questions apply to multiple spatial scales (statewide-,

regional-, or waterbody-scale) and speak to swimmable, fishable, aquatic life and tribal/cultural uses. Broad spatial and temporal coverage for California's largest lakes, reservoirs, and estuaries make OLCI satellite imagery data ideal for addressing these types of questions across a variety of temporal scales, such as annual, seasonal and/or monthly assessments. Similarly, these data can be used to assess a variety of spatial scales, ranging from assessing an individual resolvable waterbody (i.e., containing enough spatial coverage to be detected by a satellite) to assessing all resolvable waterbodies in the state. Furthermore, the Sentinel-3 mission is planned to extend to at least 2037, making this a suitable platform for an extended period of time.

The main indicators that can be derived from OLCI imagery suitable for addressing the identified management questions are total phytoplankton biomass via total *chlorophyll-a* and cyanobacterial biomass via *CIcyano*. These indicators can be applied to the four data metrics defined in this report. These metrics are bloom magnitude, bloom occurrence, bloom spatial extent, and bloom frequency, which are used to quantify near-surface, open water blooms based on their intensity and duration (magnitude), temporal patterns (frequency and occurrence) or spatial characteristics (spatial extent) within a waterbody or group of waterbodies.

The main elements used to evaluate OLCI imagery data quality are based on the Water Boards' Surface Water Ambient Monitoring Program (SWAMP) data quality assurance framework. A literature review was conducted to summarize the status of what is known and not known about OLCI data quality according to six data quality elements including the detection limit, bias, precision, representativeness, comparability, and completeness.

Although a majority of the data processing steps for OLCI imagery data are completed by the Cyanobacteria Assessment Network (CyAN) project, there are additional processing options including the application of a customized lake inventory, the further removal of potentially invalid (i.e., potentially erroneous) pixel data, or generating custom temporal composites from daily images. There are currently no universal standard practices for these additional processing options for OLCI imagery data, and different processing procedures suit different end user goals as well as their comfort with potential false positives and/or false negatives. Nevertheless, a consistent procedure with defined caveats should be developed for use of the data and for consistency between years. The report thoroughly explores these data processing options and makes recommendations for which procedures are best suited to address Water Boards management questions. These processing steps are then applied in a series of status, pattern, and trend analyses meant to provide an example of this type of assessment.

Recommended Future Actions

We identified multiple recommendations to further streamline the adoption of satellite remote sensing for FHAB assessment. These recommendations include five major types of activities:

Integrate additional satellite platforms that can expand the extent of waterbodies characterized. Higher spatial resolution platforms do exist and would allow for more waterbodies to be assessed, and for the resolution of more nearshore and narrow regions of waterbodies to be observed. Data from these platforms are valuable for future FHAB monitoring and assessment. Of greatest interest is the Multispectral Instrument (MSI) onboard the Sentinel-2 satellite constellation. Future collaboration with the CyAN project is a promising path forward to integrate data from this platform into California's growing set of remote sensing tools.

Standardize assessment of landscape and field-based indicators and metrics of FHAB drivers. To expand the utility of remotely sensed FHAB biomass assessments, we recommend assessing the available data that provides insights as to potential FHAB drivers. Drivers of FHABs consist of factors both internal and external to the waterbody. Full FHAB driver assessments require *in-situ* studies, however the pairing of remotely sensed FHAB biomass data with landscape level (i.e., land use, climate factors etc.) and field-based (i.e., water temperature, lake level, etc.) indicators and metrics of FHAB drivers complement these types of efforts and identify areas to focus efforts. Assessing satellite-based biomass dynamics in tandem with data on potential drivers can be used to identify potentially useful relationships and guide additional studies. Building this capacity includes standardizing the assessment of landscape and field-based data, identifying metrics, and workflows required to investigate the relationship between status and drivers, including the assessment of data quality.

Develop consensus on FHAB action levels for satellite remote sensing to support water quality management decisions. To improve utility of remotely sensed data for management decisions, thresholds should be defined that link satellite data (CIcyano and chlorophyll-a) to the risk of exceeding thresholds of *in-situ* photosynthetic pigment data (chlorophyll-a or phycocyanin) that impair beneficial uses in resolvable waterbodies. The infrastructure to conduct this data match up process is readily available since in recent years, approaches have been developed to streamline matching co-located *in-situ* measurements with satellite observations, making this an accessible approach.

Continue to develop quality assurance studies and documentation for application of satellite remote sensing to Water Board water quality programs. The SWAMP program should develop a generalized quality assurance project plan (QAPP) for the application of satellite remote sensing to Water Board water quality programs. This report provides the scientific review with

which to accelerate the development of that QAPP. Additionally, SWAMP should continue to validate and assess remote sensing products with available *in-situ* data to further evaluate algorithm performance. Field sampling could be aligned for dates with satellite overpasses to increase the *in-situ* data available for validation.

Enhance data communication, accessibility, visualizations, and reporting that can increase the utility of the program for the Water Boards and their partners. To expand the user community for remotely sensed FHAB data, data systems and decisions support tools are needed to encourage strong partnerships and rapid dissemination of FHAB monitoring data. To achieve this, we recommend developing the vision, including key data sources, data visualizations and graphical user interface (GUI) functionality for each type of decision support, through interactions with intended user groups. Two specific next steps are identified. These are (1) integrate satellite remote sensing into regular reporting on FHAB status and trends and (2) integrate data metrics and summaries into a FHAB satellite visualization tool to allow for functionalities such as comparison of a specific lake to a larger population of lakes and for download of data metrics into an easily digestible format such as an excel or .csv file.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Purpose of the Report

Freshwater harmful algal blooms (FHABs) are defined as an overgrowth of cyanobacteria or eukaryotic algae within a freshwater or estuarine system. FHABs have been documented throughout California's inland waters, with events being reported in lakes, rivers, streams, estuaries, and coastal confluences. FHABs can produce toxins that can harm humans, dogs, livestock, and wildlife. High biomass of both toxic and non-toxic blooms causes odor, poor aesthetics, and a cascade of ecological effects including growth of pathogenic bacteria, clogging of fish gills and benthic habitats, low dissolved oxygen concentrations, and fish kills. Additional adverse impacts of human and animal exposure to high biomass of blooms, regardless of the presence of measurable toxins, includes skin rashes, ear and eye infections, and gastrointestinal distress (Chorus and Welker, 2021).

FHABs negatively impact multiple beneficial uses in California surface waters, including drinking and municipal water, recreation, tribal and cultural uses, fishing, agriculture, and aquatic life. The degradation of these uses has broad and sustained ecosystem and economic impacts. Nutrient pollution, hydromodification, and physical habitat alteration that occur through human activities are the principal drivers, exacerbated by climate change through warming, higher CO2 levels, and changing precipitation regimes (Paerl et al., 2018). However, the extent of FHAB risks to beneficial uses is uncertain because many inland surface waters are not routinely monitored for these impacts. Due to the increasing occurrences of FHABs in California and limited resources for routine ambient monitoring, the State Water Resources Control Board's (SWRCB) Surface Water Ambient Monitoring Program (SWAMP) developed a satellite imagery tool in 2016 specifically for use in California. The satellite imagery tool was implemented as a screening tool to support prioritization for field-based surveys, such as targeting locations for HAB event response. The tool routinely acquires satellite imagery products from National Oceanic and Atmospheric Administration (NOAA) and developed as a partnership with NOAA, San Francisco Estuary Institute (SFEI), and SWAMP. To visualize the processed satellite imagery, the data is presented on a web-based tool (https://fhab.sfei.org) which provides statewide and waterbody-specific data for open-water FHABs in 255 of California's largest waterbodies.

Despite its promise for FHAB assessment, the broader applications of satellite imagery are constrained due to the limited understanding of the quality of data and lack standardized procedures for processing and analyzing satellite imagery for additional applications. Evaluating the current science of processing remote sensing data and developing applications based on

quality data promotes confidence in integrating satellite imagery data into Water Boards programs. Thereby, the development of these standardized procedures and quantifying data quality opens the door to routine use of satellite imagery in support of water quality management decisions. The FHAB Program has identified the following examples (but are not limited to): routine assessment of FHAB status and trends; integrated reporting and determination of FHAB-related water quality impairments in United States Environmental Protection Agency (USEPA) Clean Water Act Section 305(b) Integrated Reports and 303(d) listing decisions; and evaluation of management decisions and compliance with related policy.

The specific goals of this report are to: 1) Identify management questions of relevance to remote sensing of FHABs; 2) Develop standardized procedures for processing and evaluating raw remotely sensed imagery with supporting open-source data processing code; 3) Conduct a literature review to determine known data quality of these data to support Water Board decisions on quality assurance and quality control criteria; 4) Identify and document custom processes to analyze raw remotely sensed imagery to maximize resolvability of waterbodies and to address management questions; 5) Create key visualizations using these data and metrics suitable for exporting with supporting open-source code; and 6) Identify key gaps in existing FHAB Program remote sensing products and make strategic recommendations towards filling these gaps.

1.2 What California is doing about FHABs: the role of satellite remote sensing

Routine monitoring and assessment of FHABs is important to support management and stakeholder response to emerging or occurring FHABs for the protection of environmental and human health. FHABs can develop rapidly and persist for weeks or months, making them challenging to monitor due to potentially rapid changes in bloom conditions. Ambient, in-situ monitoring, therefore, is the most ideal for FHABs to protect public health, understand drivers and to understand potential mitigation options. However, routine, in-situ FHAB monitoring programs are rare in inland waters throughout the State due to limited resources as identified in the FHAB Program annual reports (California State Water Resources Control Board, 2021, 2022, 2023). Event response efforts can partially inform this gap, due to several factors including the timing of event response monitoring efforts often occurring post bloom, the turnaround time of laboratory results, and lack of follow up monitoring during an event due to limited resources and the nature of voluntary participation by local managers. These challenges are compounded due to the large number of inland waterbodies in California, some of which may be challenging to access routinely, that can experience FHAB events. Nonetheless, routine FHAB observations are needed to develop a better understanding of FHAB trends in California.

Such observations would also support the identification of systems with repeat FHAB issues to focus on where additional study of drivers and mitigation options are needed.

Assembly Bill No. 834 (AB 834) Freshwater and Estuarine Harmful Algal Bloom Program requires the SWRCB and the California Regional Water Quality Control Boards (RWQCB) to establish a Freshwater and Estuarine Harmful Algal Bloom Program to protect water quality and public health from harmful algal blooms, including conducting and supporting algal bloom field assessment and ambient monitoring at the state, regional, watershed, and site-specific waterbody scales.

With the passing of AB 834 and the establishment of a State FHABs Program, a formalized FHABs Monitoring Strategy was developed and published in 2021 (Smith et al., 2021). The document articulates the vision, framework, and strategy to develop and implement a statewide FHAB Monitoring Program. The FHAB Monitoring Strategy is centered around key management questions that are focused on immediate information needs to manage the impact of FHABs on water quality, including understanding the status and trends of FHABs, including the presence and concentrations of cyanotoxins, and the environmental drivers influencing FHAB magnitude, extent, frequency, and duration. Specific consideration in the FHABs Monitoring Strategy is given to three waterbody types: 1) lakes and reservoirs, 2) streams and rivers, and 3) coastal confluences (i.e., estuaries, coastal lagoons, etc. directly influenced by river runoff).

The FHAB Monitoring Strategy articulates the need for multiple approaches to collect ambient FHAB monitoring data to inform management decisions that ultimately lead to better protection for public health and the environment. Specifically, the strategy laid out the vision for an integrated monitoring approach with multiple elements including: 1) expanded use of remote sensing approaches that build upon the current partnership that California has formed with federal agencies; 2) a partner program that provides infrastructure to encourage FHAB monitoring by other federal, state and local agencies, tribal governments, citizen science groups, etc.; and 3) field surveys developed and managed by SWAMP or its partners. Additionally, there is need for continued and strengthened incident response efforts via synergies with ambient monitoring approaches. Meanwhile, the success of the strategy elements is reliant on data management, visualization, and decision support systems as a core part of the monitoring infrastructure necessary for managers to effectively use FHAB data for management decisions and inform timely communication to the public. Dedicated funding for the implementation of the FHAB Monitoring Strategy recommendations was not identified so the FHAB Program has focused on cost-effective approaches to implement programmatic goals to work towards the objectives mandated by AB 834 that are informed by monitoring.

Satellite remote sensing can be a cost-effective way to support several of the mandates of AB 834 and programmatic goals articulated in the FHAB Monitoring Strategy (Smith et al., 2021). Satellite remote sensing approaches can provide routine estimates of near-surface FHAB biomass estimates across broad spatial and temporal scales. Importantly, however, satellite-based biomass estimates are not meant to replace *in-situ* sampling as they do not provide critical information regarding the presence of cyanotoxins. Instead, satellite imagery data can offer a complementary assessment approach to *in-situ* monitoring programs (Shi et al., 2019) and these observations of biomass can help identify increases in biomass indicative of increased potential for cyanotoxin presence. Sensors on satellite platforms can detect light reflected from the Earth at different wavelengths that allows for the determination of cyanobacterial and algal biomass, as well as a variety of other waterbody properties such as water transparency, biological properties, hydrology, temperature, and ice cover. The overall capabilities of a given satellite mission depend on the specific sensor package(s) deployed on that satellite, and how it's operated. The spatial resolution of data is limited, in part, by the capabilities of the sensors onboard the satellite platform. The spatial and temporal resolution of satellite data products is also governed by the orbit characteristics of the specific satellite platform which influence the frequency of revisits over a given location.

1.3 Existing satellite remote sensing infrastructure for FHAB detection

1.3.1 Satellite and Sensor Infrastructure

Several international, national, and private agencies conduct satellite missions that are used to assess algal and cyanobacterial abundance across a variety of aquatic ecosystems. Federal and International agencies conducting satellite missions include the National Aeronautics and Space Administration (NASA), NOAA, and the European Space Agency (ESA). The specific capabilities of current and future satellite missions are determined by these agencies. Among these platforms, one of the most well-studied, and currently operational satellite platforms for assessing FHABs in inland waters is the Ocean and Land Color Instrument (OLCI) onboard the Sentinel-3 satellite constellation. Sentinel-3A was launched by the ESA in February 2016, and Sentinel-3B was launched in April 2018 (Clerc et al., 2020). Sentinel-3A provided observations every 2-3 days beginning 2016 until the inclusion of Sentinel-3B data in May 2018 after which observations were available near-daily (Schaeffer et al., 2022). The lifetime of these satellite platforms is currently estimated at ten years. Additional Sentinel-3 satellites are included in the ESA Sentinel-3 program with a Sentinel-3C launch projected for 2026 and a Sentinel-3D launch projected for 2028, for a currently planned life of this program of at least twenty years. The OLCI sensor has a 300 meter by 300-meter (~22 acres) pixel resolution for a given area of the

state, though factors including cloud cover, sun glint, snow, and ice can limit image availability during satellite overpass (see section 2.3 for full discussion of observation limitations).

OLCI can detect light reflected from the Earth at multiple wavelengths. This allows for the estimation of cyanobacterial and algal abundance in individual satellite pixels using spectral shape algorithms for specific spectral bands in the upper portion of the water column. This includes differentiating the unique spectral characteristics of cyanobacterial and algal biomass based on the unique pigments (e.g., phycocyanin) present in cyanobacteria (see Chapter 3 for an extended description of the algorithm). Prior to OLCI, the Medium Resolution Imaging Spectrometer (MERIS), onboard the Envisat satellite platform, had similar sensor characteristics and was used to study FHABs until it became non-operational in 2012. Currently, approaches to detect FHABs, including the specific detection of cyanobacteria, from imagery data collected by these platforms are well studied for use in lentic waterbodies such as lakes and reservoirs (Clark et al., 2017; Coffer et al., 2020, 2021a, 2021b; Urquhart et al., 2017; Wynne et al., 2010). Alongside these technological advancements to detect FHABs formed by cyanobacteria, researchers also developed algorithms to estimate chlorophyll-a (Chl-a; a proxy for overall algal and cyanobacterial biomass or density). Because cyanobacterial dominance and cyanotoxin concentrations are strongly correlated with chl-a (US EPA 2021), chl-a becomes a useful proxy for eutrophication and an added indicator of potential cyanoHAB risk, even when cyanobacterial blooms are not detected. Thus, given the paucity of *in-situ* data currently available for the assessment of FHABs in California's waterbodies, OLCI imagery offers a powerful tool to begin to assess the status and trends of eutrophication in general and cyanoHAB events in large lakes and reservoirs. As more *in-situ* data becomes available, these data can supplement the remotely sensed data to improve the accuracy and completeness of these assessments.

1.3.2 Data delivery and visualization infrastructure

Satellite remote sensing is already being implemented as an approach to guide monitoring FHABs within California. Since 2016, California has routinely acquired remotely sensed cyanobacterial concentration estimates from NOAA via OLCI with historic satellite data from MERIS going back to 2002. Through this strategic investment, these data are provided through a California FHAB satellite portal hosted by SFEI (https://fhab.sfei.org/). The purpose of this tool, as it currently exists, is to provide early warning for managers to mobilize field crews and to encourage verification and surface water sampling to inform public health advisories. It is particularly useful to prioritize locations for event response field assessments in large waterbodies. A recent application of OLCI imagery for event response occurred in the San Francisco Bay estuary in 2022 and 2023. Large blooms of a marine algae called *Heterosigma akashiwo* occurred in the estuary, resulting in widespread fish kills and ecosystem impacts. Chl-

a estimates from OLCI imagery helped to guide and focus monitoring efforts during these events by providing synoptic views of the Bay to identify where algal biomass was the greatest. Imagery data also provided an important temporal time series on the development of the bloom as it spread through the estuary.

The current CA FHAB satellite portal was not specifically designed to assess trends in cyanobacterial and algal abundances over time, or to support water quality management decisions recommended by the FHAB Program, such as impairment listing, TMDL development, or compliance with biostimulatory water quality objectives. Therefore, these investments, while successful for event response, have not yet resulted in extensive applications to address FHAB management questions or actions.

Several efforts at the federal level are promising to help support the expanded use of satellite remote sensing data to address a broader set of management questions about FHABs. The Cyanobacteria Assessment Network (CyAN) project is a collaborative initiative across the USEPA, NASA, NOAA, the United States Geological Survey (USGS), and, as of 2023, also the United States Army Corps of Engineers (USACE) to monitor and assess cyanobacteria in resolvable waterbodies across the United States. The CyAN project conducts initial data processing of raw OLCI/MERIS data into a standardized, publicly accessible output for use in cyanobacterial assessment. These are served via the NASA website at: https://oceancolor.gsfc.nasa.gov/about/projects/cyan/.

In addition to serving data products, the CyAN project has developed recommended approaches for processing and analyzing OLCI imagery data for identifying and tracking cyanobacterial blooms on the national scale in resolvable waterbodies. Multiple peer reviewed papers have published approaches for assessing cyanobacterial bloom occurrence, spatial extent, magnitude, frequency, and trends from imagery data. As a part of these efforts, some initial assessments of national trends in FHAB occurrence, frequency, extent, and magnitude have been conducted (Coffer et al. 2020, 2021a, 2021b; Urquhart et al. 2017; Mishra et al. 2019; Clark et al. 2017; Shaeffer et al. 2022).

1.4 Current Needs

The FHAB Monitoring Strategy identified that strategic investments to strengthen California's use of remote sensing would be beneficial to the FHABs Monitoring Program and other related water quality management programs. These tools provide cost-effective and complementary information to field-based assessments of FHAB status, trends, and drivers. In this report, we focus on four concepts that were identified by the FHAB Program that would allow remotely sensed imagery to be more completely utilized by relevant Water Boards Program areas: 1) Development of standardized procedures to routinely process and evaluate best practices,

including routine raw data with known quality; 2) document data quality control to support the FHAB Program's development of a specific quality assurance plan; 3) provide data communication, accessibility, visualizations, and reporting that can increase the utility of these data for the Water Boards and their partners and; 4) develop technological improvements that can greatly expand the number and completeness of resolvable waterbodies for evaluations.

The Water Boards prioritize documentation of data quality so that they can make management decisions with confidence. Therefore, assessment of the quality of data derived from remotely sensed imagery is critical to support any downstream uses of the data. Standardized analyses of remotely sensed imagery are important for consistency and interpretation of these data over time. Standardization of data of any type helps to eliminate discrepancies, ambiguity, and interpretation by end users and elicits confidence in the use of the data for management decisions. For the purposes of the FHAB Monitoring Program, the goal of using remotely sensed imagery is to extract meaningful data about FHABs in California's inland waters. To accomplish this goal, a variety of data processing steps are needed to transform imagery into usable data metrics for this goal. Standardizing these data processing steps as best practices means that the caveats and limitations of the data can be clearly identified and articulated when communicating any downstream analyses and that the processing can be readily repeated. Standardized imagery processing also means that the processed data will be in a standardized format, making long term data management, assessment, and storage easier.

1.5 Report Goals and Organization

The organization of this report is as follows.

In **Chapter 1**, there is an introduction that describes the purpose and goals of this report, the current approach for monitoring and managing FHABs and how satellite remote sensing can support these efforts, existing satellite remote sensing infrastructure for FHABs, report organization and complimentary documents, and the approach to developing this report.

Chapter 2 articulates the focal management questions that guided the data quality assessment, data processing and data analysis development efforts.

Several cyanobacteria specific indices and data metrics can be calculated using MERIS, and now OLCI imagery. These are described in detail in **Chapter 3**.

A description of data processing steps, options and recommendations are described at length in **Chapter 4** of this report. A step-by-step guide for processing including the use of open-source R code for analysis is described in the corresponding **Guidance Document**.

In **Chapter 5**, a description of data analysis methods to support answering management questions related to monitoring.

A data quality literature review was conducted and is presented in **Chapter 6** of this report. An additional complimentary effort for satellite remote sensing field validation was also conducted as a part of this work. The results of those efforts are presented in the corresponding **Field Validation Report,** but key findings of that effort are presented in this chapter.

Chapter 7 presents a vision for future applications and recommendations for future work.

1.6 Approach to Developing this Report

The overarching approach to developing this report and recommendations consisted of the use of a Technical Advisory Committee (TAC). The TAC consisted of two working groups. The first was a technical working group, consisting of national experts in FHAB ecology and satellite remote sensing. The technical working group provided technical feedback on data quality assessment, data processing and data analysis elements, discussed various approaches and their inherent advantages and disadvantages, and helped provide consensus recommendations, where possible. The second working group within the TAC was an end user working group consisting of key technical staff from State and Regional Water Boards and other state agencies and tribes, and community scientist groups, who provided broad perspectives on the utility of, and application needs for integrating satellite remote sensing data products into programs, policies, regulations, and monitoring efforts.

FHAB Water Boards staff proposed a set of key management questions (described in detail in **Chapter 2**) that reflected agency priority information needs and are consistent with the information needs of other potential partners from the federal to local level. These management questions were initially vetted by the TAC, then were refined by the subgroups to discuss explicit recommendations for satellite remote sensing data that could (or could not) answer those questions.

All management questions and work products focus on filling in key gaps for the application of OLCI imagery. The initial focus was limited to OLCI data products from onboard the Sentinel-3 satellite mission for several reasons. The first is that the OLCI data products can provide estimates of both overall algal biomass via Chl-a concentration estimates, as well as cyanobacterial abundance estimates via the CIcyano index. This is due to the specific wavelength bands available via OLCI data products. Additionally, large investments by both California and federal agencies have already been made to do much of the data processing, data hosting and serving of OLCI data products, making use of these data much more accessible to a variety of users. The major drawback of OLCI data products is a larger pixel size (300 m),

which limits the number of inland waterbodies resolvable using this platform. The large pixel size also limits the ability to resolve nearshore and narrow regions of large waterbodies. Nonetheless, the benefits of this platform outweigh these limitations for assessing FHABs in the surface waters of resolvable waterbodies.

Future efforts will be focused on imagery that has finer spatial resolution, and the framework and analyses described in this report are envisioned to be scalable to these types of higher spatial resolution imagery data products. Of greatest interest is the Multispectral Instrument (MSI) onboard the Sentinel-2 satellite constellation. This sensor offers a 20-meter spatial resolution which will allow for a majority of waterbodies in California (roughly 14,500 waterbodies are estimated to be resolvable statewide) to be monitored remotely with far greater spatial resolution of areas near shorelines and beaches, which are known to concentrate FHABs and thereby strengthen the detection of FHABs near recreational areas. MSI generates imagery with a different spectral signature; therefore, these products cannot be used to calculate a CIcyano value. Methods have been developed to use MSI imagery to estimate Chl-a concentrations, which could provide data about overall algal biomass like OLCI, but a cyanobacteria specific index equivalent to the CIcyano index is not currently available. Nonetheless, MSI and other high resolution satellite imagery products offer promising additions to FHAB monitoring. One of the major limitations for including MSI data products in this current effort is that pre-processed data products from Sentinel-2 are not currently available on a routine basis. Due to the increased spatial resolution, a significantly larger amount of data is generated by this platform compared to OLCI imagery. Significant investments would be needed to manage and use this data. The CyAN project has kicked off efforts in 2024 to build similar infrastructure as currently available for OLCI data products. Thus, future use of Sentinel-2 and other higher resolution data products were kept in mind throughout the development of this report, and many of the recommendations for data processing and analysis in this report are envisioned to be scalable to these future high resolution data products.

2. MANAGEMENT QUESTIONS ADDRESSED VIA SATELLITE REMOTE SENSING

2.1 Management Questions

The FHAB Program, along with key management, governmental and stakeholder entities, have a common set of information needs related to FHABs and how they impact the beneficial uses of California's inland waterbodies. The FHAB Monitoring Strategy document outlines these questions extensively across multiple spatial scales and a wide variety of monitoring approaches (Smith et al., 2021, Chapter 2). On a broad level, these questions focus on status, trends, environmental drivers, and FHAB incident response. The questions apply to multiple spatial scales (statewide-, regional-, or waterbody-scale) and speak to swimmable, fishable, aquatic life and tribal/cultural uses.

In this present work, this framework was applied to satellite remote sensing imagery with a focus on OLCI imagery, although applicable to potential other satellite platforms, and efforts focused on developing the approaches to address questions related to these themes. The questions guiding this current effort are described in Table 2.1.

Table 2.1: A summary of the four overarching management question categories that guided the development of OLCI data processing and analysis workflows.

Category	Questions
Status	What is the overall magnitude and frequency of FHABs in OLCI resolvable waterbodies in CA?
	How often do blooms occur in OLCI resolvable waterbodies in CA?
	What is the spatial extent of OLCI resolvable water area in CA that experiences a bloom?
Trends	How are bloom occurrence, magnitude, extent, and frequency changing over time?
Drivers	What are the environmental factors that are associated with bloom occurrence and magnitude?
Incident Response	What is the ongoing risk in the waterbodies in my region or watersheds with past FHAB event reports? How can remotely sensed data be used to inform incident response?

2.2 Remotely Sensed FHAB Indicators and Linkages to Impact Pathways

The main indicators that can be derived from OLCI imagery are biomass based and thus represent at least one risk pathway to each of the beneficial use categories (*see Table 2.2 of Smith et al., 2021*). The impacts of biomass across these priority beneficial uses are briefly described here (*see Appendix 2 of Smith et al., 2021 for an expanded discussion of FHAB impacts*).

Effects of excessive biomass on aquatic life, terrestrial wildlife, and livestock occur through multiple pathways. As the extent, frequency, and magnitude of algal blooms begin to increase, marked changes to primary producer biomass and community structure fundamentally restructure food webs that support invertebrates, fish, birds, amphibians, and other wildlife. The biomass of nutrient tolerant, opportunistic epiphytic and drift micro- and macroalgae and phytoplankton can increase under these scenarios and these species can dominate an ecosystem. High biomass can also cause shifts in dissolved oxygen and pH levels that negatively impact aquatic life. Cyanotoxins may also be more prevalent and can cause a variety of chronic and acute ecotoxicological impacts (Mehinto et al., 2021).

Contact (REC1) and non-contact (REC2) recreation are impacted by FHABs through a variety of pathways. Visual scums, poor water clarity, the “pea green soup” of high biomass blooms, and abundant cyanobacterial mat growth in shallow water are all examples of aesthetic impairment of waterbodies. Increased organic matter accumulation associated with high biomass blooms, coupled with low dissolved oxygen concentration, can cause a proliferation of heterotrophic bacteria, some of which may be pathogenic to aquatic organisms and humans (NRC 2000). Direct impacts to human health and domestic animal health from cyanotoxin exposure, including but not limited to skin rashes, ear and eye infections, neurological symptoms, and gastrointestinal distress, are significant risks that are increased with higher cyanobacterial biomass (Puschner et al. 2008; Stewart et al. 2008; Backer et al. 2013).

Fishable beneficial uses are impacted through several pathways. High biomass FHAB events can result in hypoxia, shellfish disease, fish kills, and the mortality of other aquatic species (Glibert et al. 2002) that ultimately reduces abundance and biodiversity of aquatic and terrestrial wildlife (e.g., salmonids, crabs, bivalves, et al. sportfish) associated with the beneficial uses of aquaculture (AQUA), shellfish harvesting (SHELL), and commercial and sportfishing (COMM). Additionally, bioaccumulation in fish and shellfish tissues of cyanotoxins may make them risky to consume (Miller et al. 2010; Kudela 2011), or taste and odor compounds can cause taste issues (Burr et al. 2012; Howgate 2004; Robin et al. 2006), respectively.

High biomass blooms are problematic for raw drinking water source protection and municipal waters for several reasons. Beyond the increased risk of toxins and taste and odor compounds, increased dissolved organic carbon (DOC) results from “leaky” algal blooms. Higher DOC levels increase the amount and costs of disinfectants required to achieve disinfection goals. DOC, algal metabolites, and other decomposition products, when present in raw water and chlorinated or brominated by treatment processes, can produce trihalomethanes (THM), which include several known and suspected carcinogens. (USEPA 2000; Graham et al. 1998; Plummer and Edzwald 2001). High biomass blooms also can impede municipal or industrial water intakes.

Tribal uses of waterbodies are impacted by high biomass FHABs through a variety of pathways. Tribal and cultural uses of waterbodies are very specific to each tribe, and it is difficult to summarize the diversity of uses simply. Each tribe in the State is a sovereign nation with distinctive cultural practices and traditional uses of waterways. These uses of the water are often extensive and involve significantly more and different types of exposure than recreational uses. Depending on the specific tribe and tradition impacts from uses may include those already described above or may include unique impacts due to a range of distinctive exposure pathways.

2.2.1 Chlorophyll-a

Bulk measures of photosynthetic abundance and biomass provide an understanding of the trophic status of a waterbody. One of the most common indicators of photosynthetic algal abundance (including cyanobacteria) is Chl-a. Concentrations of Chl-a in surface water can be optically determined via satellite imagery. The amount of Chl-a in water changes the waterbody’s absorption and reflection of light. An algorithm (algorithms described in **Chapter 3**) that uses specific satellite wavelength bands is then to estimate the near-surface Chl-a concentration from the spectral data. Concentrations of Chl-a in surface water can also be determined by field-based procedures, each with their specific procedures, equipment, and quality control; these procedures include laboratory analysis of collected water samples or field-based measurements using hand-held instruments.

2.2.2 Cyanobacterial biomass

Cyanobacterial biomass in a waterbody can provide a relative estimation of the risk that a waterbody may experience impacts related to FHAB events. Assessment of bulk cyanobacterial abundance is useful as an initial assessment of hazard as increased cyanobacterial biomass is related to an increased potential of cyanotoxin presence (US EPA 2021). Cyanobacteria have Chl-a but also have an accessory pigment called phycocyanin which is a unique accessory pigment that gives cyanobacteria distinctive spectral qualities. Cyanobacterial biomass can be differentiated using a distinct wavelength signature that allows for the differentiation of

cyanobacterial biomass from other algae and optically active matter (see **Chapter 3** for extended explanation). A recent study using OLCI and MERIS imagery demonstrated this linkage by comparing CIcyano derived estimates of cyanobacteria to *in-situ* presence of microcystins in eleven US states, including California. Bloom detection accuracy between same day *in-situ* sample and satellite imagery match ups was 84%. This same study projected an accuracy range of 77%-85% with a 95% confidence interval based on a bootstrapping uncertainty analysis (Mishra et al., 2021).

2.3 Strengths and Weaknesses of Remote Sensing for Addressing Management Questions

Satellite imagery data is different from other data collected in the field in that no physical access to the field is necessary, providing information about a wide area of waterbodies, including hard-to-access waterbodies where there is little or no existing data. Because the data is collected in a routine way, it is well suited to automated and scalable analysis of one or many waterbodies, decreasing the amount of effort to analyze the data. In addition to the high spatial coverage, satellite imagery data also has high temporal coverage, with nearly daily imagery collected by Sentinel-3 satellite overpasses. This data is also complementary to field-based assessments of status and trends and provides one of the most consistent and longest time series of any type of water quality data to fill in data gaps related to FHAB assessments.

The synoptic view that satellite imagery provides makes it particularly useful for assessing the status and trends of FHABs, which can be ephemeral or episodic in nature. This synoptic view also allows for comparing California data to national or global trends in FHABs. Furthermore, utilizing satellite imagery data can be very cost effective compared to the resource and effort required for field monitoring programs. California is estimated to save $120,000 per year using satellite imagery to monitor 20% of lakes resolvable by Sentinel-3 (Papenfus et al 2020). This estimate is grossly underestimated based on California's current monitoring costs. Since field-based monitoring programs in the state are very limited in scale and frequency of assessments due to low resources, satellite imagery is critical to begin assessing the impacts of the resolvable lakes, reservoirs, and channels. While satellite imagery cannot detect if cyanotoxins are present, it can detect presence of cyanobacteria and estimate cyanobacterial concentrations. The magnitude of cyanobacterial biomass is an indicator of decreased water quality and greater densities, detected in open water of resolvable lakes, can increase the potential threat of exposure by humans and animals thereby prioritizing waterbodies for field-based cyanotoxin measurements.

Despite these strengths, some key limitations and caveats of satellite imagery data exist:

1. **Satellite imagery data is limited to optically detectable parameters**. Satellite data can only measure parameters that are detectable optically, such as differences in wavelengths that can be used to infer photosynthetic pigment concentrations and thus only provide estimates of algal and cyanobacterial concentrations. This is an important distinction in the case of FHABs, since cyanotoxins cannot be detected by satellite imagery. Thus, satellite imagery is not currently applied to inform California state voluntary recreational health advisories, which are based on cyanotoxin concentrations and *visual* confirmation of a bloom. Furthermore, biomass levels need to be high enough concentrations to be detectable via satellite sensors, therefore more disperse or lower biomass blooms may not be detected. Similarly, while cyanobacteria and other types of algae can be differentiated via optical properties, it is not currently possible to differentiate the different types of cyanobacteria present since there are not distinctive enough optical properties. Differentiating cyanobacterial taxa based on their spectral properties, however, is an area of current research (Legleiter et al., 2022). This also means that other relevant parameters like nutrient concentrations and water flow are not detectable via imagery data.

2. **Satellite imagery data is limited to surface measurements.** Satellite imagery can only measure the water surface where light can penetrate. The depth coverage of this measurement is variable based on the turbidity or cell concentration of the water. Many cyanobacteria can regulate their buoyancy and often form dense populations near the water surface, which makes surface measurements by satellites favorable. However, blooms that form subsurface or on the benthos are unlikely to be detected via satellite imagery, unless waters are exceptionally clear or shallow. This limitation may also result in potential false negatives or underestimate biomass for planktonic cyanobacterial species that regulate their buoyancy and are submerged at the time of the satellite overpass. Since California does not have a comprehensive ambient monitoring program for FHABs, the abundance of cyanobacteria species that form blooms lower in the water column and on the benthos is unknown, so the severity of the limitation is unknown.

3. **Satellite imagery data is limited to open water.** Satellite imagery integrates the optical properties over the resolvable area of the sensor (e.g., the pixel). This means when pixels contain both land and water (a mixed pixel), the signal of the water and land are integrated. This integration may result in erroneous results depending on the proportion of land and water in a mixed pixel. This is particularly the case when the shoreline area has a substantial amount of vegetation, which contains Chl-a, that is challenging to disentangle from any Chl-a present in the nearby water. Thus, mixed pixels may need to be removed during the processing of satellite imagery data to reduce this uncertainty (see more in **Chapter 4** about options for removal). With the removal of potential mixed

pixels, most often near the edge and shore of waterbodies, detections in the open water pixels are an elevated concern for potential denser cyanobacteria abundance near the edges, particularly driven by wind and water current. Thus, open water detections may underestimate or not detect blooms along the edges that are common locations of beaches, boat launches, and other more publicly accessible areas. Measurements of optical properties through satellite remote sensing are also subjected to possible top-of-atmosphere radiance contamination from neighboring surfaces and/or aerosols with different reflectance. This is termed 'adjacency effect' and can be an additional source of error for mixed pixels. Thus, removal of mixed pixels is one conservative approach to minimize adjacency effects, although these can also be addressed using various algorithms.

4. **Satellite imagery data from OLCI is limited to larger and wider waterbodies.** The mixed pixel paired with the larger pixel size of OLCI imagery, limits the size and shape of waterbodies that can be detected as the waterbody must have at least one complete open water pixel. The waterbody size requirement is often higher than 90,000 m^2 (300 m x 300 m) for OLCI data because it is rare that a pixel (which is a fixed grid) will neatly encompass water only for small waterbodies. Narrow waterbodies are also often not resolvable for the same reason. Fortunately, the open water regions of over 200 of the largest and widest lakes, reservoirs, and channels are resolvable so many of the critical drinking water reservoirs and water storage basins have satellite imagery data to assess. Smaller waterbodies that are not resolvable via OLCI (estimated to be at least 14,000 waterbodies) will be resolvable in the future and add to the dataset available for assessment with the growing science, launching of new satellites, and use of satellite imagery with greater resolution.

5. **Environmental conditions can cause data gaps in satellite imagery data**. Clouds, sun glint, snow, ice, and atmospheric conditions can cause interfere with the collection of satellite imagery. If these conditions are present during the satellite overpass, then algal and cyanobacterial abundance data cannot be derived from those images. This can result in extended periods of lost data if these conditions are persistent.

These strengths and weaknesses result in some key caveats to each of the specific management question categories presented in Table 2.1.

2.4 Management Questions about Status and Trends

Status and Trends management questions: What is the overall magnitude and frequency of FHABs in OLCI resolvable waterbodies in CA? How often do blooms occur in OLCI resolvable waterbodies in CA? What is the spatial extent of OLCI resolvable water area in CA that experiences a bloom? How are bloom occurrence, magnitude, extent, and frequency changing over time?

Strengths: OLCI satellite imagery data is well poised to address the identified management questions about both the status and trends of cyanobacterial and algal biomass. Broad spatial and temporal coverage make OLCI satellite imagery data ideal for addressing these types of questions across a variety of temporal scales, such as annual, seasonal and/or monthly status and trend assessments (Table 2.1). Similarly, these data can be used to assess status and trends across a variety of spatial scales, ranging from assessing an individual resolvable waterbody to assessing all resolvable waterbodies in the state. Furthermore, while trend assessments are currently limited to 2016 to present, the ESA Sentinel-3 program plans extend to at least 2037, thus the ability to assess trends will only increase over time.

Weaknesses: These questions can be answered in waterbodies if they are large enough for at least one open water OLCI pixel (e.g., 300m x 300m) to be detected without interference from land or mixed pixels. Thus, primarily only large lakes and reservoirs would be detectable using these data. Intermittent gaps in assessment of status and trends are also possible due to environmental conditions but overall, the number of observations possible via satellite remote sensing are still greater than those via current *in-situ* monitoring efforts. Notably, in California, cloudy and icy conditions are most prevalent in the winter and early spring seasons in the northern areas of the state, and coastal fog is common in some areas such as the San Francisco Bay in the summer. Blooms of cyanobacteria have been observed during these periods thus these conditions could result in some underestimation of blooms in these areas during this part of the year.

Applications: Biomass estimates are useful for identifying when, where, and how often cyanobacterial blooms occur (see section 2.2 for impacts of high biomass on beneficial uses). Also, biomass estimates can provide an indication of where cyanotoxins might be present, since increasing cyanobacterial biomass is correlated to increasing cyanotoxin concentrations. It is important to note, however, that this is a risk-based relationship and correlations between biomass and cyanotoxin presence can be imprecise and vary based by waterbody (see Howard et al., 2021 and Smith et al., 2023 for examples from California lakes based on *in-situ* data).

2.5 Management Questions about Drivers

Drivers management question: What are the environmental factors that are associated with bloom occurrence and magnitude?

Strengths: Remote sensing data can be a powerful tool for risk assessment by providing a synoptic dataset of bloom biomass indicators for the development of statistical associations between increased instances of blooms with a variety of potential environmental drivers. While OLCI imagery does not provide environmental data directly, many driver data sets are publicly available such as land use, precipitation, or hydrodynamics (e.g., lake level, lake discharge, etc.), that can be used to identify correlative relationships with environmental conditions for future study.

Weaknesses: Driver assessments via satellite imagery are limited to assessing the drivers of increased biomass and cannot provide direct information related to drivers of cyanotoxin production or concentrations.

Applications: The driver relationships that can be derived via remote sensing data correlative are risk-based relationships. These relationships are for prioritizing resolvable waterbodies for future *in-situ* monitoring or specialized studies to identify drivers more precisely. These observations can also be applied for assessing risk of blooms in lakes that are not resolvable via OLCI imagery but share similar predictor characteristics.

2.6 Management Questions about Incident Response

Incident response management questions: What is the ongoing risk in the waterbodies in my region or watersheds with past FHAB event reports? How can remotely sensed data be used to inform incident response?

Strengths: Satellite imagery can capture a synoptic image of a waterbody, which can help identify the location of a bloom and estimate the bloom extent. Field based sampling, particularly in large waterbodies, is very spatially limited and thus can potentially miss bloom activity.

Weaknesses: The limit of detection for cyanobacterial specific biomass estimates is high compared to other *in-situ* monitoring methods. Thus, *in-situ* monitoring methods, if conducted, can detect biomass at earlier stages than satellite imagery.

Applications: Satellite imagery can provide near real-time indication of where a bloom is forming in near-surface, open water areas of resolvable waterbodies and can also guide field-

based event response efforts in large waterbodies, such as in the San Francisco Bay (see **Chapter 1**) and large reservoirs in the California State Water Project. Also, considering the longer temporal extent of cyanobacterial blooms often observed in California lakes and reservoirs, the near real-time biomass detections inform when and how often follow-up monitoring for an event should occur to assess public health impacts until the bloom dissipates. Satellite imagery detections showing re-occurring cyanobacterial biomass across seasons and years supports prioritization of resources for proactive *in-situ* monitoring programs to assess public health risk.

3. ASSESSMENT METHODOLOGY: INDICATORS, METRICS, AND SPATIOTEMPORAL REPORTING UNITS FOR FHABS ASSESSMENT

Indicators and ***metrics*** describe what information derived from satellite imagery data will be used to address management questions. An ***indicator*** is the type of measurements, while ***metric*** is the precise measurement methodology and resulting values from measurements over time and space. In this chapter, the basic description of each indicator derived from OLCI imagery data will be discussed followed by a synthesis of the relevant metrics using these data. The same metrics described below can be applied to other satellite platforms such as the MSI onboard Sentinel-2 as part of future efforts.

3.1 Indicators

3.1.1 Cyanobacteria Index

The CI value is a proxy of cyanobacteria specific Chl-a absorption and estimates the cyanobacterial biomass using a distinct spectral shape signature that allows for the differentiation of cyanobacterial biomass from other eukaryotic algae and reflective matter present in water (Wynne et al., 2008, 2010). This CI value was calculated using established spectral shape algorithms that utilize the spectral shape centered at 681 nm (SS(681); λ^- = 665 nm; λ^+ = 709 nm). A negative shape occurs at 681 nm as a function of low fluorescence by cyanobacteria, and high backscatter which overwhelms absorption at 709 nm, and has been suggested to be the result of the structure of the cells gas vacuoles (Wynne et al., 2008). As such, this index can flag other blooms from non-cyanobacteria phytoplankton such as diatoms and chlorophytes. To reduce the occurrence of non-cyanobacteria phytoplankton as false positives of cyanobacteria, the CI was further refined to include the spectral shape centered at 665 nm (SS(665); λ^- = 620 nm; λ^+ = 781 nm) that is sensitive to phycocyanin absorption at 681 nm. This new CIcyano value allowed the separation of cyanobacteria blooms from other blooms by the following rule:

If SS(665) < 0, cyanobacteria is not present and CIcyano = 0

If SS(665) > 0, cyanobacteria is present and CIcyano = CI

Note that CIcyano has also been called CImulti, referring to its use of multiple shape algorithms. A detailed description of the CIcyano calculations can be found in Lunetta et al. (2015) and a description of the development of CI and later CIcyano can be found in Coffer et al. (2020).

3.1.2 Digital number

Satellite data downloaded from the CyAN project is presented as digital numbers (DN) for each pixel, where values range from 0 to 255, and values ranging from 1 – 253 indicate cyanobacteria detection. Using the conversion equation of:

$$CIcyano = 10^{(DN*0.011714-4.1870866)}$$

The lowest DN of 0 indicates that cyanobacteria are below the detectable limit of the satellite sensor (but not necessarily that cyanobacteria are completely absent). DNs of 254 and 255 are data flags, with a DN value of 254 indicating land, and a DN value of 255 indicating no data was available, due to factors such as cloud cover, sun glint, or other data quality flags.

3.1.3 Cell abundances

At least two studies have been published to examine the relationship between *in-situ* cell abundances and imagery-based estimates of cyanobacterial density. The first study was conducted in western Lake Erie that used a small sample set (n = 12) to compare cell count estimates against co-located CI values derived from MERIS imagery (Wynne et al., 2010). This relationship was defined by the following equation:

$$Cell\ abundance\ \left(\frac{cells}{mL}\right) = 6.9\text{x}10^{7} * CI + 1.8\text{x}10^{5}$$

A later study by Lunetta et al. (2015) compared CIcyano values to *in-situ* cell count observations (n = 2068) in lakes from Ohio, New England (RI, MA, NH, ME, NY, and VT), and Florida. CIcyano values were converted to cell abundances estimates using the following equation:

$$Cyanobacterial\ cell\ abundance\ \left(\frac{cells}{mL}\right) = CIcyano * 10^{8}$$

This study reported a good correspondence between *in-situ* cyanobacterial cell counts and imagery cell count observations was reported with an R^2 value of 0.87 being reported for the regression analysis comparing measures. An important caveat to this finding is that ultimately a much smaller subset of the *in-situ* samples (n = 579) was compared to imagery data, with 72% of the samples being removed from the regression analysis comparing the two data sources, which improved the model fit. In particular, the comparisons indicated that the equation was robust for low (10,000 – 109,000 cells/mL) and very high (>1,000,000 cells/mL) cell abundance but did not perform as well for an intermediate range (110,000 – 1,000,000 cells/ml). The lower performance of the intermediate range may be attributed to the lower amount of data available (much of the cell count data in this category was dropped from the analysis) as well as

uncertainties or inconsistencies in the cell count dataset. An additional confounding factor is that the CI and later the CIcyano algorithms were developed from observations of Lake Erie which is primarily dominated by *Microcystis* spp., whereas the lakes used for comparison to CIcyano values in Lunetta et al. (2015) were often mixed assemblages of cyanobacterial taxa (though extensive documentation of community composition was not presented), thus the performance of the cell conversion is less well understood in mixed cyanobacterial communities, which are often documented in California waterbodies.

A more conservative approach to estimate cyanobacterial abundances is to convert CIcyano to categorical abundance ranges with the goal of adding some context to the less easily interpreted CIcyano value. This approach is described in Clark et al. (2017) and Coffer et al. (2021b). In this approach, each resolvable pixel was categorized as having cyanobacteria above or below the sensor detection limit, which is estimated to be between 10,000 cells/mL - 20,000 cells/mL (Coffer et al., 2021a). Then, if cyanobacteria were detected, the pixel was categorized as having low, moderate, or high human health risk based on cell concentration ranges defined by World Health Organizations (WHO) guidance in 1999. WHO associates bloom abundances of ≤20,000 cells/mL with a low risk of adverse health impacts. A moderate risk of adverse health impacts is defined as between 20,000 cells/mL - 100,000 cells/mL and high risk is defined as **≥**100,000 cells/mL (Chorus and Bartram, 1999).

FHAB Monitoring Strategy discourages the use of specific cell densities estimated from OLCI satellite imagery (Smith et al., 2021). While published methods to estimate cyanobacterial cell abundances exist, there are a multitude of associated caveats with the approach that lead to this recommendation, which is continued in this report. Due to the uncertainty associated with cell abundance conversions and limitations of the currently published studies on the topic, cell abundance conversions are generally not recommended for application without additional validation and study.

3.1.4 Total and Cyanobacterial Chlorophyll-a

Chl-a is a proxy for total algal biomass, which includes both cyanobacterial and non-cyanobacterial algae. This is distinct from the specific measurement of cyanobacterial biomass that is achieved using CIcyano. However, estimations of Chl-a concentrations can be derived from remotely sensed imagery data using multiple approaches. It can be detected directly using directly using specific satellite wavelength bands or determined through conversions of CIcyano to an estimated Chl-a concentration. A key difference exists between these two approaches is that direct detection of Chl-a using specific sensor wavelength bands is estimated a Chl-a concentration that reflects all algal biomass (e.g., both cyanobacterial and non-cyanobacterial algae) which we describe in this report as total Chl-a. In comparison, conversion of Cycyano

values to an estimated Chl-a concentration is a reflection of cyanobacterial Chl-a concentrations due to the specificity of the CIcyano algorithm to cyanobacterial biomass.

A variety of algorithms exist for estimating total Chl-a. Here, the Red Edge 2010 (RE10) algorithm presented by Wynne et al (2022) is used. This matches the approaches on the California FHAB satellite tool. The RE10 algorithm examines the ratio of a red-edge band (665 nm) to the near infrared band (709 nm). They are atmospherically corrected by taking the difference of the 885 nm band, which represents an assumption of coarse-mode maritime aerosol with dark water. This is first calculated as the R2 ratio using the following equation where ρ_s is the top-of-atmosphere (TOA) reflectance:

$$\mathrm{R2} = \frac{\rho_s(709) - \rho_s(885)}{\rho_s(665) - \rho_s(885)}$$

Next, the RE10 regional adjustment algorithm is applied as follows:

$$RE10 = [35.75 \times R2 - 14.30]^{1.124}$$

This is then reported as a concentration of micrograms per liter and was field tested using samples collected in Chesapeake Bay (Wynne et al., 2022).

Because total chl-a is proposed as a key indicator in the interpretation of California's narrative biostimulatory water quality objective (Sutula et al. 2025), it is useful to retain and track this indicator alongside cyanobacterial-specific indicators and metrics.

3.1.5 Conversion to Cyanobacterial Chlorophyll-a

In addition to calculating Chl-a directly from imagery data, there are two published equations suitable for converting CIcyano values to estimated cyanobacterial Chl-a concentrations. These conversions represent a Chl-a concentration that is attributable to cyanobacteria:

$$Cyanobacterial\ Chl = 4050\ (\pm 271) * CI + (\pm 3)$$ Tomlinson et al. 2016

$$Cyanobacterial\ Chl = 6620\ (\pm 646) * CIcyano - 3.1\ (\pm 5.2)$$ Seegers et al. 2021

One of advantages of converting CIcyano values to Chl-a concentration equivalencies is to make the reporting of CIcyano more intuitive for those less familiar with remote sensing of cyanoHABs and the scaling of DN. Updated WHO guidelines utilize Chl-a observations in a risk based alert framework for waterbodies that are dominated by cyanobacteria (Chorus and Welker, 2021). The WHO recommends a Chl-a concentrations >12 µg/L as an Alert Level 1 threshold in waters dominated by cyanobacteria (Chorus and Welker, 2021). Using a conversion of CIcyano values to a Chl-a value fits well in the WHO framework, since CIcyano values

specifically indicate cyanobacterial presence. The WHO threshold of 12 µg/L is useful to apply to define a bloom alert level of cyanobacteria and is recommended for integration into several of the metric calculations described in the section below (see Section 3.3). A Chl-a concentration of 12 µg/L is approximately equivalent to a DN of 132 using the Seegers et al. (2021) equation. This approach mirrors the methods implemented by Schaeffer et al. (2024), where the WHO Alert Level 1 threshold is utilized in a satellite imagery-based bloom forecasting framework.

The conversion equation described in Tomlinson et al. (2016) equation was developed using field radiometry CI and Chl-a data from Florida and may therefore be more specific to that region. Even though CI is used to develop the equation, Florida waterbodies tend to be dominated by cyanobacteria, thus the CI is comparable to CIcyano in those waterbodies. The range of Chl-a in the dataset used to develop the equation was 16 – 115 µg/L, thus this relationship with CI at Chl-a concentrations < 16 µg/L is not confirmed. Seegers et al. (2021) analyzed this equation and found a $bias_{log}$ of 1.33, indicating that the Chl-a value estimated from the equation is on average 33% higher than the dataset used to evaluate the equation. In addition, Seegers et al. (2021) found that the equation worked best with Chl-a concentrations > 20 µg/L.

The Seegers et al. (2021) equation was developed using extracted Chl-a concentrations from the USGS CyAN Field Integrated Exploratory Lakes Database. The authors reported a positive $bias_{log}$ of 1.11 (11%) and mean absolute error (MAE_{log}) of 1.6 (60%) for the equation when compared to *in-situ* observations. This represents a slight improvement on performance compared to the conversion equation reported in Tomlinson et al. (2016). Like the Tomlinson et al. (2016) equation, the performance of the Seegers et al. (2021) equation is poorest at oligotrophic and mesotrophic waters with lower Chl-a concentrations (< 7 µg/L).

3.2 Recommended spatial and temporal reporting units

The synoptic nature of satellite imagery allows for the application of management questions across a range of spatial scales. Imagery data is readily parsed into a variety of spatial units. This report focuses on statewide and waterbody specific spatial units. These were identified as the most suitable for the number of waterbodies resolvable via OLCI imagery due to the overall number of waterbodies resolvable across California. In total, 238 large lakes, reservoirs and estuaries are resolvable in the state (see **Chapter 5** for more details on lake inventory development). While this is a major increase in the number of routinely assessable waterbodies in California compared to *in-situ* monitoring efforts, it is still a fraction of the total number of waterbodies statewide. FHAB Monitoring Strategy estimated that more than 14,000 lakes,

ponds, and reservoirs 1 hectare or larger exist statewide (Smith et al., 2021). Finer scale regional assessments, such as assessing waterbodies within the bounds of a RWQCB, are also possible using OLCI imagery data. The important caveat is that the number of resolvable waterbodies is highly variable from region to region. Thus, while regional assessments are certainly possible, it is recommended that any comparative regional analyses with OLCI imagery data be approached with caution due to the limited number of resolvable waterbodies in some regions.

Sentinel-3 overpasses occur every ~1-2 days, resulting in multiple images available per week for a given area of the state. The CyAN project data portal makes daily imagery available as well as 7-day maximum composites. For the purposes of this project, 7-day maximum composite images were identified as ideal for metric calculations since it smooths potential data gaps from environmental conditions (e.g., clouds, ice, sun glint, etc.) that may occur within a week and requires less computational power than daily imagery to process data and metrics. Metrics can be calculated on a weekly, seasonal, or annual basis to adequately address specific management questions.

3.3 Metrics

Four data metrics integral to comprehensive status and trend assessments have been developed by the CyAN project in recent years and these are bloom magnitude, bloom occurrence, bloom spatial extent and bloom frequency. Also, these metrics are also consistent with the ambient status and trends assessments conducted by SWAMP for non-HAB water quality monitoring. These metrics focus on quantifying blooms based on their intensity and duration (magnitude), temporal patterns (frequency and occurrence) or spatial characteristics (spatial extent) within the resolvable portion of waterbody or group of waterbodies. These calculations can be made for either total Chl-a, CIcyano or cyanobacterial Chl-a, depending on the applications and needs of the end user. Throughout this report, we apply these metrics with CIcyano and cyanobacterial Chl-a as examples, but these examples could also be applied using total Chl-a.

3.3.1 Magnitude

Bloom magnitude indicates the intensity of the cyanobacteria (or total algal) presence over a defined period of time (Figure 3.1). Mishra et al. (2019) defined this as the spatiotemporal mean of CIcyano within a waterbody for the week, season, or year, normalized by surface area:

$$Bloom\ magnitude = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{T}\sum_{t=1}^{T}\sum_{p=1}^{P} CIcyano_{p,t,m}$$

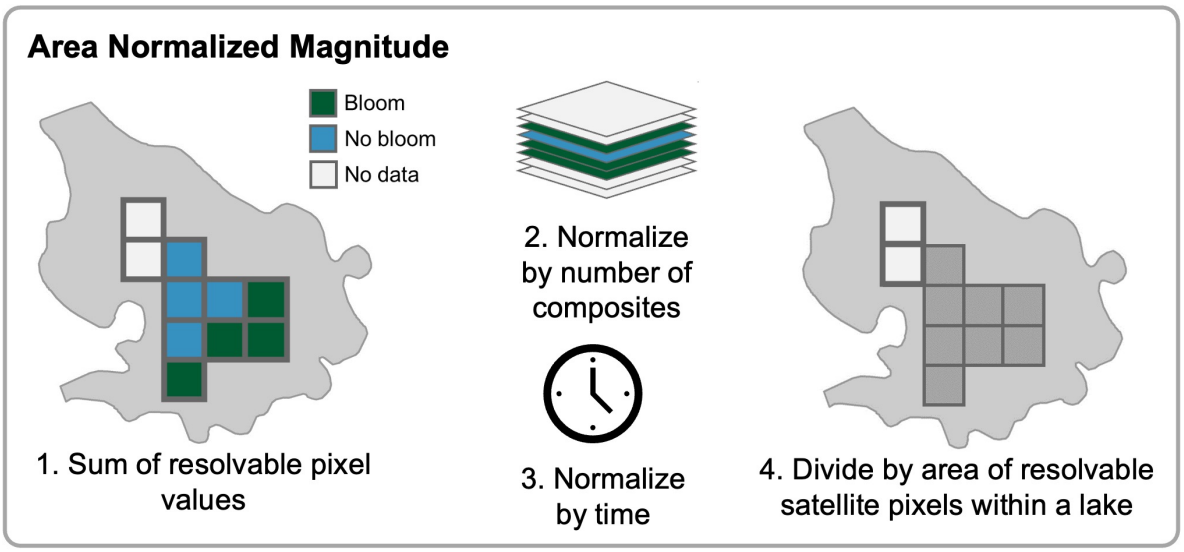

Figure 3.1: Conceptual diagram illustrating the area normalized bloom magnitude metric. Diagram based on the formulas and concepts described in Mishra et al. (2019).

Where M is the number of weeks or months in the time being assessed; T is the number of composites per study period (e.g., four 7-day composites per month); P is the number of valid pixels for the waterbody. The area normalized bloom magnitude is then calculated from bloom magnitude as:

$$Area\ normalized\ bloom\ magnitude = \frac{Bloom\ magnitude\ (CIcyano)}{Waterbody\ surface\ area\ (km^2)}$$

The authors used this metric to quantify annual and seasonal cyanobacteria bloom magnitude for lakes in Florida and Ohio during 2003 – 2011 using the functionally similar MERIS sensor onboard the Envisat satellite.

3.3.2 Occurrence

Cyanobacterial bloom occurrence summarizes pixel level data to determine if a bloom occurred in a single waterbody on a weekly basis based on an area threshold and cyanobacterial density threshold (Figure 3.2). This metric was used to assess the overall seasonal patterns of blooms within the state. Bloom occurrence was calculated following the methods described in Coffer et al. (2020) as the percentage of lakes experiencing a cyanobacterial bloom above each threshold on a weekly basis. The occurrence metric defines a bloom based on a minimum spatial bloom area (e.g., a percentage of the detectable pixels of a waterbody) and CIcyano threshold. We used a spatial threshold of 10%, meaning that for a lake to be considered as having a bloom, at least 10% of the total number of detectable pixels in a waterbody must be at or above each cyanobacteria abundance threshold. This spatial threshold was recommended by Coffer et al. (2020) as it reduces the variance in occurrence across differently sized waterbodies and across

seasons. Two thresholds for bloom intensity were used. The threshold for a bloom is cyanobacterial detection via remote sensing (DN > 0). The threshold for a bloom alert is defined as ≥12 µg/L cyanobacterial Chl-a using the conversion factor described in Seegers et al. (2021; DN ≥ 132). Using these criteria, bloom occurrence can be calculated from each weekly image composite into counts of waterbodies experiencing a bloom.

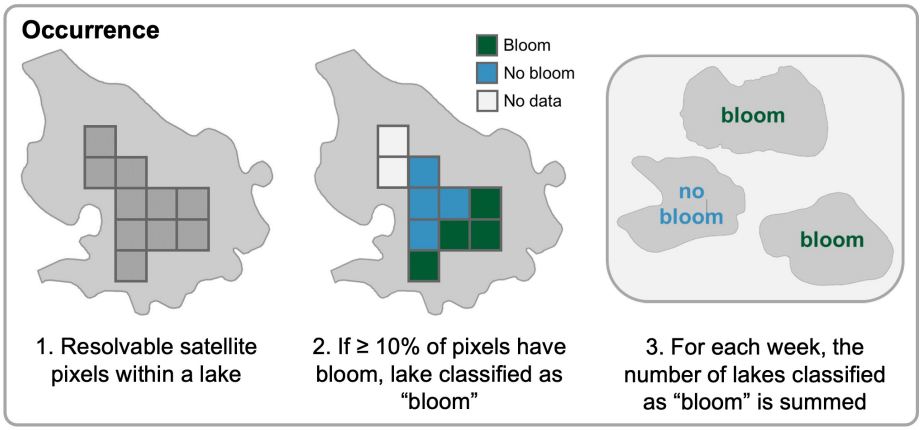

Figure 3.2: Conceptual diagram illustrating the bloom occurrence bloom metric. Diagram based on the formulas and concepts described in Coffer et al. (2020).

3.3.3 Spatial Extent

Spatial extent summarizes the percentage of resolvable waterbody area with detectable cyanobacteria presence (Figure 3.3). This metric was first used to assess the temporal changes of cyanobacteria bloom extent in Florida, Ohio, and California (Urquhart et al., 2017), before efforts were expanded across CONUS (Schaeffer et al., 2022). Spatial extent is calculated by:

$$Spatial\ extent\ (\%) = \frac{number\ of\ pixels\ with\ detectable\ CI_{cyano}}{number\ of\ valid\ pixels}\ \text{x}\ 100\%$$

Weekly spatial extent can be calculated for individual lakes, as well as summarized for a region (e.g., State). Pixels that are designated as land or no data due to cloud coverage, and pixels that are removed from any masking processes (Details in **Chapter 5**) are not counted towards the calculation, thus the number of valid pixels can be variable over time.

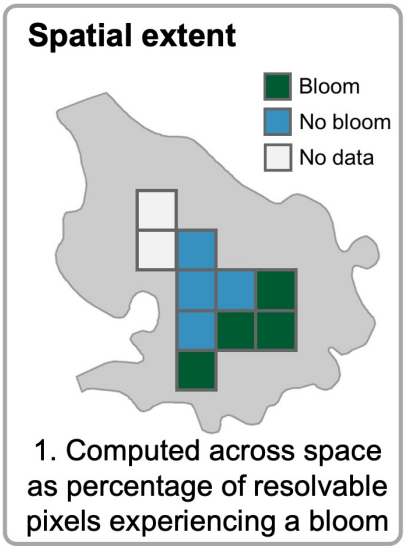

Figure 3.3: Conceptual diagram illustrating the bloom spatial metric. Diagram adapted from Schaeffer et al. (2022).

3.3.4 Frequency

The cyanobacteria bloom frequency metric is defined for each satellite pixel as the percentage of valid satellite images for that pixel throughout the year exhibiting cyanobacteria above each cyanobacteria abundance threshold (Figure 3.4). The development and refinement of this metric is described in Clark et al. (2017) and Coffer et al. (2021b). As with the occurrence metric, the cyanobacterial Chl-a bloom alert threshold can be applied to integrate an estimation of bloom density into the temporal frequency metric. Temporal frequency was computed as:

$$Annual\ pixel\ scale\ frequency = 100\% \times \frac{n\ of\ pixels\ at\ identified\ threshold}{n\ of\ all\ valid\ pixels\ in\ the\ year} \quad (1)$$

Individual lake frequencies were calculated by averaging the annual frequencies for all detectable pixels within each lake according to equation (1):

$$Annual\ lake\ specific\ frequency = \frac{\sum annual\ pixel\ scale\ frequency\ in\ lake}{n\ of\ resolvable\ pixels\ in\ lake} \quad (2)$$

Statewide frequency calculations were made by averaging the annual frequency of all the detectable water pixels calculated via equation (1) within California state borders according to equation (2):

$$Annual\ statewide\ frequency = \frac{\sum annual\ pixel\ scale\ frequency\ in\ CA}{n\ of\ total\ resolvable\ pixels\ in\ CA} \quad (3)$$

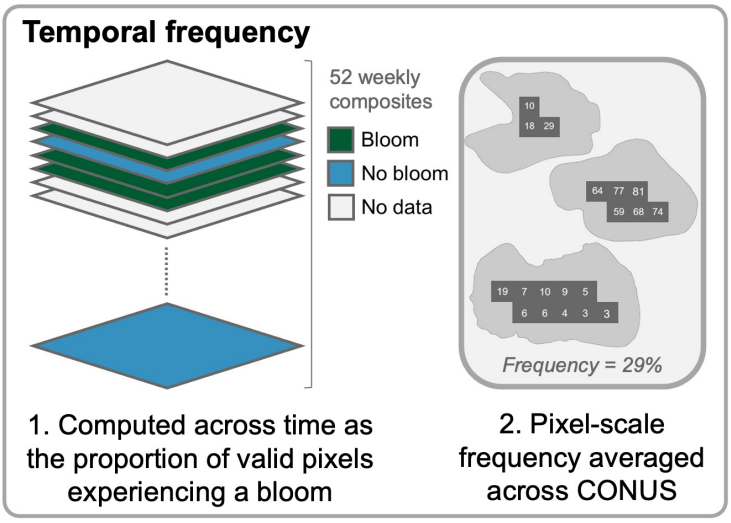

Figure 3.4: Conceptual diagram illustrating the bloom frequency metric. Diagram adapted from Coffer et al. (2021b).

4. BEST PRACTICES FOR FHAB ASSESSMENT DATA PROCESSING

4.1 Comparison of different processing options

The CIcyano product is processed from raw imagery data and made available by the CyAN project. Standardized processing is conducted by the CyAN project for these products to relieve intensive data processing burdens from the end user. In brief, imagery data are geolocated, converted to an Albers Equal Area projection, and corrected for top-of-atmosphere reflectance to remove the spectral contribution of Rayleigh scattering. The CIcyano values are calculated, and data flags are applied to pixels containing clouds and land (including probably mixed land-water pixels) using standardized algorithms. Imagery data are available publicly as a 7-day maximum composite or are available as daily images.

After these data are downloaded, there are a variety of additional data processing steps that can be conducted depending on the goals of the user. These additional processing options can include the application of a customized lake inventory, the further removal of potentially invalid (i.e., potentially erroneous) pixel data, or generating custom temporal composites from daily images (Table 4.1). There are currently no universal standard practices for these additional processing options for OLCI imagery data, and different processing procedures suit different end user goals as well as their comfort with potential false positives and/or false negatives. Nevertheless, a consistent procedure with defined caveats should be developed for use of the data and for consistency between years. The data processing options explored in this chapter are described here along with final selections for procedures to best address the management questions articulated in **Chapter 2**.

4.2 Inventory of resolvable lakes

Data processing using satellite imagery begins by defining the inventory of waterbodies that will be analyzed. Currently, a variety of shapefiles exist including shapefiles published by the CyAN project. Users can also choose the waterbodies of interest for processing through the creation of a custom shapefile. This step serves to 1) focus the data processing and analysis to only waterbodies of interest (identified as polygons in the shapefile), thereby significantly reducing computing power requirements, and 2) provide metadata information (e.g., Lake name, COMID, and coordinates) in summaries generated during data analysis.

Table 4.1: An overview of various data processing options available for different aspects of satellite imagery processing. The ultimate selections and recommendations made after assessing these options are bolded in the table.

Description	Options available
Inventory of resolvable lakes (Section 4.2)	• Use existing shapefile products • **Develop a California custom shapefile**
Removal of mixed land-water pixels (Section 4.3)	• No removal • Removal by static TIF indicating invalid pixels • **Automated removal of boundary pixels** • Generation of buffer zone and automated removal of pixels not contained within buffer zone
Removal of snow and ice pixels (Section 4.4)	• No additional removal • **Removal using 4 km ice resolution**
Temporal compositing of the data (Section 4.5)	• **Temporal composites** vs. daily imagery data • **7 days** vs 10 days for temporal composites • Median vs **Maximum** value for composites

The CyAN project developed a national inventory of resolvable lakes within the contiguous United States that is available as a product on their project website. The shapefile related to this inventory contains 2,321 waterbody polygons across the nation. The construction utilized the National Hydrography Dataset Plus version 2.0 (NHD Plus) and includes waterbodies ranging in surface area from $7.5x10^5$ m^2 to over $4.0x10^9$ m^2. (Urquhart and Schaeffer, 2020). A total of 88 of the polygons in this inventory are located in California. A small number of these 88 polygons represent multiple segments within a waterbody, so when these are combined, there are a total of 83 waterbodies in California within this inventory.

Since the inventory developed by CyAN contained a small percentage of all waterbodies in California, we developed a custom inventory to expand the number of resolvable waterbodies available within the state. This custom California waterbodies shapefile includes 238 waterbodies for the use of OLCI imagery data processing, a 170% increase in the number of resolvable waterbodies included in the CyAN inventory. This shapefile was generated by first combining the polygons from the NHD (version NHD_H_California_State_GDB) and NHD plus (version NHDPlusV21_CA_18_NHDSnapshot_05) shapefiles, which consisted of more than 14,000 waterbodies across the state. Seventeen additional polygons for the San Francisco Bay and Delta region were copied from a shapefile provided by SFEI that was developed in collaboration with the Water Boards (Figure 4.1; FHAB_Lake_BufferedBoundaries_2019_05_16.shp). An initial size filter of $2.7x10^5$ m^2

(equivalent of three 300 m x 300 m OLCI pixels) was then applied to reduce the number of waterbodies to ~2,000, and these remaining waterbodies were manually inspected with ArcGIS Pro World Imagery and Google Earth to ensure they contained water and were cross-compared to the common grid used in CyAN data to ensure they contained at least one complete 300 m x 300 m OLCI pixel within the bounds of the waterbody polygon (Figure 4.2).

An attribute called 'Water Type' was also assigned during this manual inspection process to categorize waterbodies into 3 categories:

- Lakes and reservoirs (total = 210). Within this group, additional designations were made:
    - Lake: waterbodies identified as a natural lake
    - Reservoir: waterbodies that were identified as created by the presence of a dam or other artificial boundary based on researching the waterbody
    - Agricultural: to denote type converted waterbodies within agricultural land-uses (AG) with straightened (e.g., rectangular) rather than natural shorelines (total = 10).
- Tidal: Selected estuaries sufficiently large to be resolvable. Specifically, this included San Francisco Bay and San Joaquin Delta and large, episodically open meromictic coastal lagoons (total = 28). Other large estuaries have the potential to be resolvable but were not specially included in the workflow due to large marine influences (e.g., San Diego and Mission Bays, Tomales Bay, Humboldt Bay, etc.).

The shapefile was also manually curated to merge polygons that had initially split a waterbody (e.g., Clear Lake in Figure 4.3A), but retained different sections for the San Francisco Bay polygons (based on management of the system). Additional features and/or shapefile boundaries within the waterbody that may interfere with data processing (e.g., Mono Lake in Figure 4.3B) but were not visible in true color imagery were also removed.

Preliminary processing of data from 2017 – 2023 was conducted with the draft inventory to provide additional information on the number of resolvable pixels for each waterbody. These results were used to determine if additional polygon curation was needed. Preliminary processing identified six waterbodies had 0 resolvable pixels throughout the 7 years and those waterbodies were removed from the shapefile (not included in the 238 waterbodies in the final shapefile). An additional 47 waterbodies had only 1 – 2 potentially resolvable pixels, or a high number of weeks with 0 resolvable pixels (51 – 90% of the 366 weeks). These 47 waterbodies were categorized with an attribute called 'Suitability for trend analysis' as 'limited' to indicate these waterbodies have limited number of routinely resolvable pixels and should not be included in overall summaries and trend analyses due to data quality and availability concerns

(Table 4.2). The remaining waterbodies have 3 or more potentially resolvable pixels, and preliminary data analysis indicated they had <50% of 366 weeks with 0 resolvable pixels. These waterbodies were assigned the attribute “suitable”.

Figure 4.1: Overview of the 17 polygons for the San Francisco Bay and Delta region. These polygons were imported from a shapefile obtained from SFEI and the Water Boards.

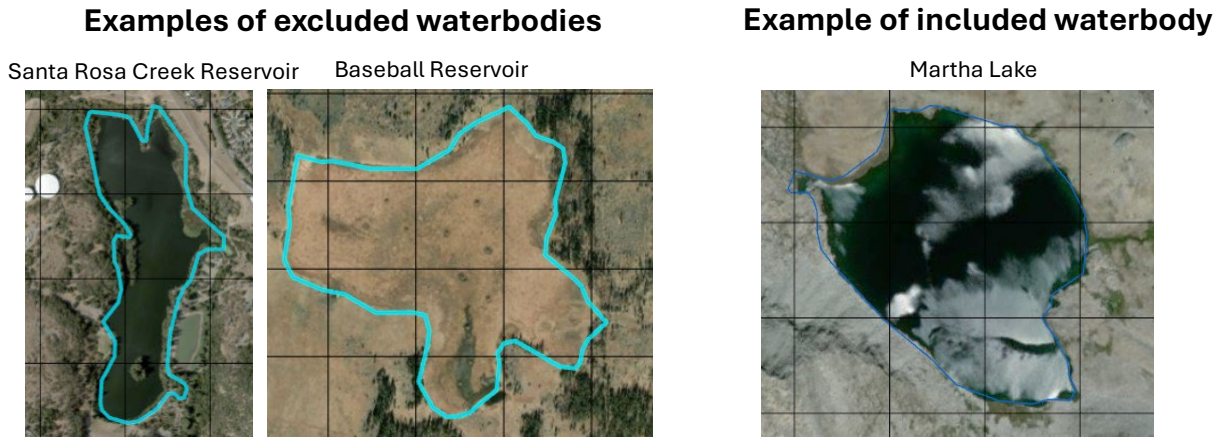

Figure 4.2: Examples of excluded and included waterbodies for custom shapefile construction. Santa Rosa Creek Reservoir was excluded as it does not contain a complete pixel, while Baseball Reservoir was excluded as it appeared to be dried up. Black squares indicate OLCI common grid pixels. Martha Lake was included in the inventory, while Santa Rosa Creek Reservoir and Baseball Reservoir were excluded from the inventory. True color basemap imagery courtesy of ArcGIS Pro World Imagery (Map Version 2023).

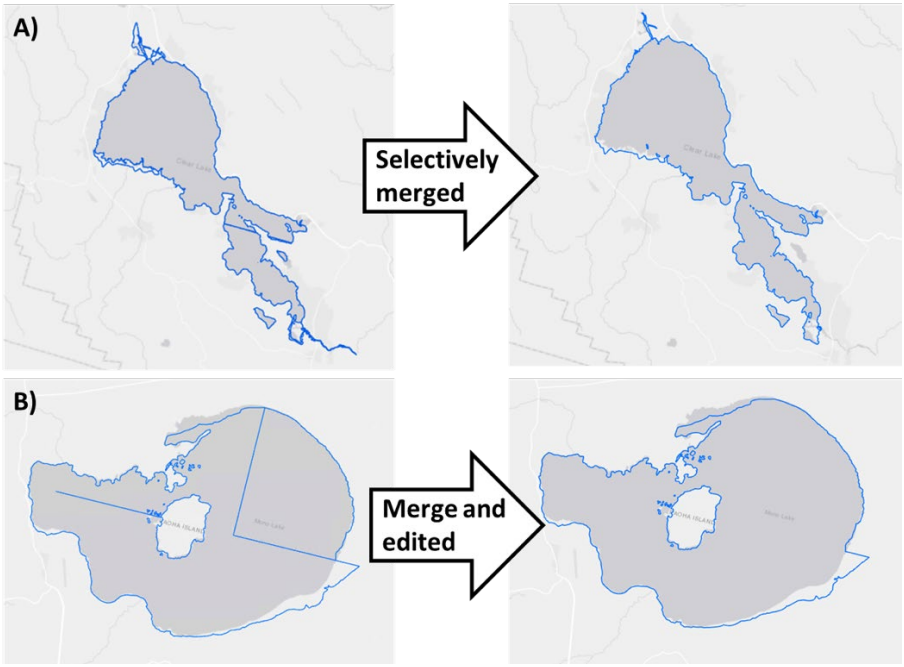

Figure 4.3: Demonstration of the manual curation of polygons for the CA custom shapefile. A) Merging of selected split sections of Clear Lake. B) Merging split sections of Mono Lake, and removing a line in the polygon that does not appear to be any physical feature of the lake as indicated through current and historical Google Earth satellite images.

Table 4.2: Polygon quality criteria. Criteria were based on an assessment of pixel number and resolvability from 7 years of imagery data between 2017-2023.

Polygon Usage Category	Inclusion Criteria	Number of waterbodies in this category
Limited	1 – 2 pixels OR 51-90% of Weeks with 0 pixels	47
Suitable	3 or more pixels AND 0-50% of Weeks with 0 pixels (> 1 resolvable pixel > 75% of the 6 years)	191

4.3 Removal of mixed land-water pixels

4.3.1 Mixed pixel flagging

The processing of satellite imagery data commonly includes steps that identify pixels that are unsuitable for further processing due to concerns about their accuracy or quality. Data downloaded from the CyAN project data portal are already flagged for pixels obscured by cloud or sun glint, and these pixels are given a DN of 255 to indicate 'No Data'. Pixels that are dominated by land are also able to be flagged, and the CyAN project flags these with a DN of 254. These invalid pixels are easily removed during processing since they are already flagged with specific DN values.

While there is an algorithm used to flag land and pixels containing a mix of land and water ('mixed land-water pixels' or 'mixed pixels' for short), some mixed pixels may not be detected and flagged by the algorithm. These mixed pixels may contain variable ratios of land and water, and currently it is not well understood how much land needs to be present in a mixed pixel for it to be flagged as land. Mixed pixels that are dominated mostly with water may also yield usable data, but it is dependent on a variety of factors including how much of the pixel area includes water, the magnitude of reflectance from nearby land (e.g., adjacency effect), presence of shoreline vegetation, or submerged vegetation in shallow waters. These factors may lead to erroneous interpretation of light spectra in these pixels as the presence of cyanobacteria. Determining the presence of potentially mixed pixels relies on identifying these pixels using another reference such as looking at true color imagery or creating a shoreline buffer area to remove these pixels. An added challenge in identifying and removing these pixels from further analysis is variable water levels that occur over time due to a variety of factors including water discharges, flooding, or drought.

4.3.2 Impact of mixed pixel flagging on resulting bloom metrics

We compared the results of a variety of approaches for addressing mixed pixels to determine how they might influence further analyses and related conclusions. The goal of this exercise was to better understand if mixed pixel flagging approaches might yield significantly different answers about the number of bloom occurrences, spatial extent, temporal frequency, and bloom magnitude of cyanobacterial blooms in California waterbodies within a year and on an annual basis.

Imagery data from the CyAN project data portal (version 5 7D max composites 2017 - 2023) was processed using a modified version of the shapefile provided by the CyAN project website (https://oceancolor.gsfc.nasa.gov/images/about/MERIS_OLCI_Lakes.zip) that included 83

waterbodies in California. The period of 2017-2023 encompasses years with high levels of precipitation (2017, 2023) and drought periods (2020-mid 2022), thus a variety of water levels are represented across this period. This shapefile was used instead of the custom-made shapefile described above because it pairs with a static TIF file that was manually curated by Urquhart and Schaeffer (2020) to flag mixed land-water pixels by referencing true color satellite imagery. The original shapefile had 88 polygons for California, but manual inspections of each polygon revealed some polygons to be segments of the same waterbody, so some polygons were merged (similar to the processing shown in Figure 4.3). The paired static TIF was also updated manually to account for modification to the polygons, as well as recent changes in water levels and water boundary location by referencing satellite imagery from ArcGIS (Map Version 2023; source dates from 2020 – 2022) and historical imagery (2017 – 2023) from Google Earth. The use of this static TIF for flagging mixed pixels was compared to skipping additional mixed pixel removal, automatically removing pixels at the boundary of the polygon, or automated removal with a 150 m or 300 m buffer zone from the original polygon boundary. The inclusion of a buffer zone is a more conservative (e.g., avoiding pixels with greater potential for false positives) option than simple boundary removal as it removes all pixels that are within or intersecting with the 150 m or 300 m buffer of the polygon boundary. For consistency, these data were all further processed by removing ice pixels by referencing ice and snow coverage from the National Snow and Ice Data Center (4 km resolution) following Urquhart and Schaeffer (2020).

In summary, the sets of data processed for this comparison included:

1. No removal of mixed pixels
2. Automated removal of intersecting polygon border pixels
3. Removal of mixed pixels using a static TIF following Urqhuart and Schaeffer (2020)
4. Automated removal of pixels within a 150 m buffer zone of the polygon border
5. Automated removal of pixels within a 300 m buffer zone of the polygon border

Figure 4.4 illustrates the differences between the different processing results using Pyramid Lake as an example.

The removal of mixed pixels dramatically reduced the number of pixels analyzed (Figure 4.5A) but did not necessarily affect the average bloom frequency or the annual total number of bloom occurrences on an annual basis (Figure 4.5B - D). Not surprisingly, the use of a 300 m buffer zone for mixed pixel removal had the lowest number of pixels analyzed each year and reduced the number of waterbodies that could be resolved (i.e., with at least 1 potentially

resolvable pixel) by 14. This may explain the lower count of weekly bloom occurrences for this processing method compared to the other methods (gold bars in Figure 4.5B). The average annual bloom detection frequency (defined as DN > 0) and average annual bloom alert frequency (defined as DN $\geq$ 132) followed the same trend of 'No mixed pixel removal' giving the highest value, followed by 'Automated removal', 'Automated removal with 150 m buffer zone', and 'Automated removal with 300 m buffer zone'. The results of mixed pixel removal using a curated static TIF was a bit more variable depending on years and the parameter, but in general the results for that method are most comparable to results of using a 150 m buffer zone.

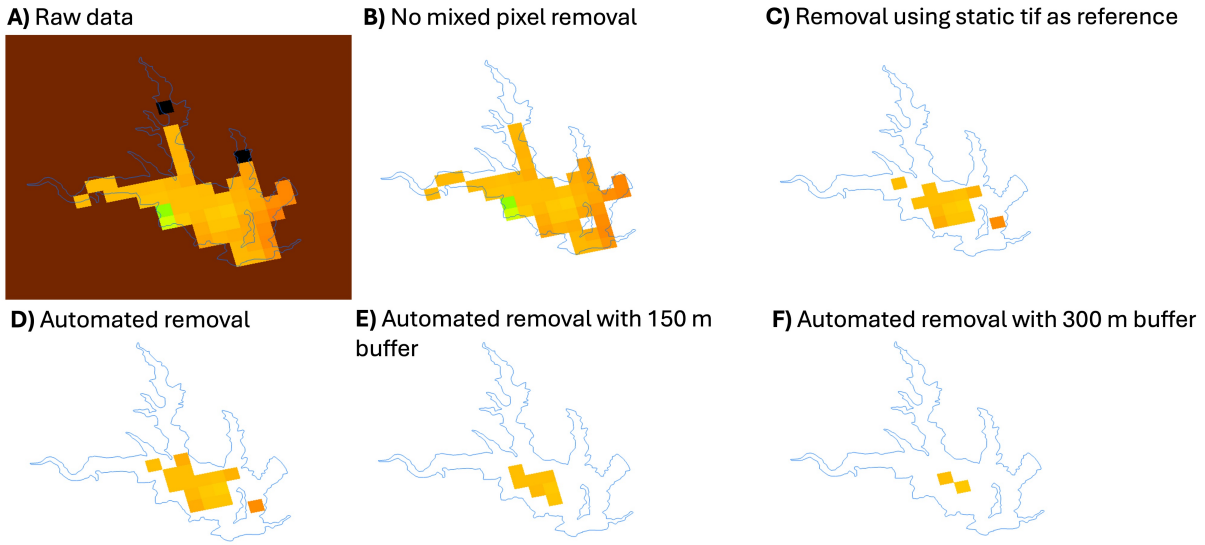

Figure 4.4: Different processing method for mixed land-water pixels using Pyramid Lake in week 30 of 2020 as an example. A) Raw data from CyAN. Brown pixels in the raw data indicate land pixels flagged by CyAN project data processing procedures (DN = 254) and black pixels indicate 'No data' pixels flagged by the CyAN project (DN = 255). B) Removal of land and 'No data' pixels only. C) Removal of potential mixed land-water pixels by using a static TIF for reference. D) Automated removal of polygon border pixels (i.e., Any pixels touching the border of the polygon was removed). E) Automated removal of pixels within a 150 m buffer zone of polygon border. F) Automated removal of pixels within a 300 m buffer zone of polygon border.

A more temporally resolved examination of weekly state-wide bloom spatial extent (percentage of pixels with DN > 0 against all resolvable pixels) and number of weekly bloom occurrences in 2017 (high rain year) and 2022 (drought impacted year) revealed a similar trend of a higher value for 'No mixed pixel removal', followed by 'Automated removal', 'Automated removal with 150 m buffer zone', and 'Automated removal with 300 m buffer zone'. Occasionally, using the 300 m buffer zone led to a higher bloom extent (e.g., 2 gold peaks between week 10 – 20 for

2022 in Figure 4.6B), but this is likely due to occasions when shoreline pixels, which are excluded by the 300 m buffer zone, has pixels with lower DN compared to the center of the waterbody, leading to a higher extent as a lower total number of pixels is used to calculate the percentage. This is atypical for cyanobacteria blooms as they typically accumulate at the shoreline (Gons et al., 2005). Bloom magnitude follows the same pattern of a higher value for 'No mixed pixel removal', followed by 'Automated removal', 'Automated removal with 150 m buffer zone', and 'Automated removal with 300 m buffer zone'. The results of mixed pixel removal using a curated static TIF was again more variable depending on years and parameter. Overall, the different processing methods led to comparable bloom extent, weekly bloom occurrence, and bloom magnitude except when a 300 m buffer zone was used.

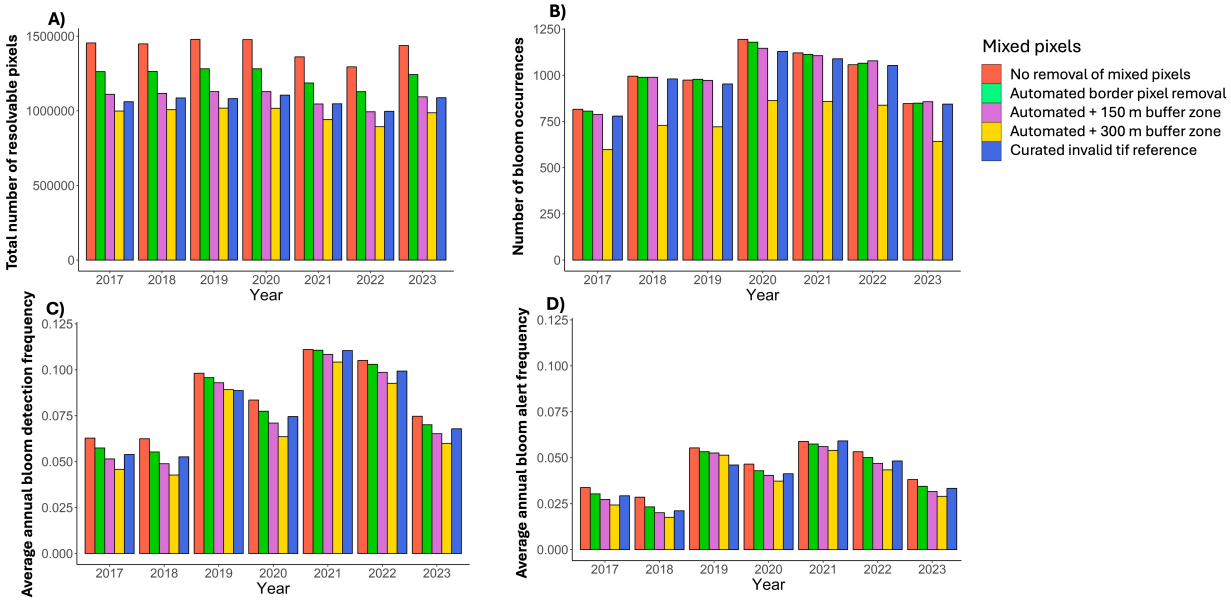

Figure 4.5: Comparison of different mixed pixel inclusion or exclusion methods on A) total number of pixels processed in the year; B) total number of bloom detection occurrences (DN > 0 and spatial extent > 10%); C) average annual bloom detection frequency (DN > 0); and D) average annual frequency of bloom alerts (DN ≥ 132) for 83 waterbodies. Bloom frequency is a proportion.

4.4 Removal of snow and ice pixels

4.4.1 Snow and ice flagging

Ice/snow coverage can also affect the optical signals used to calculate CIcyano, often resulting in false positives. It is difficult to identify snow/ice pixels directly with OLCI imagery data and flag them in the same way that clouds or glint are detected. Pixels that potentially contain ice

and/or snow undergo an initial flagging process via a NOAA algorithm, but the algorithm is often not full effective at identifying these pixels. Thus, additional flagging of ice pixels should be considered by using other references dataset of the presence of ice/snow (Urquhart and Schaeffer, 2020). Urquhart and Schaeffer (2020) released a workflow and recommendations for removing pixels potentially contaminated by snow and ice using data from the National Snow and Ice Data Center as a reference database.

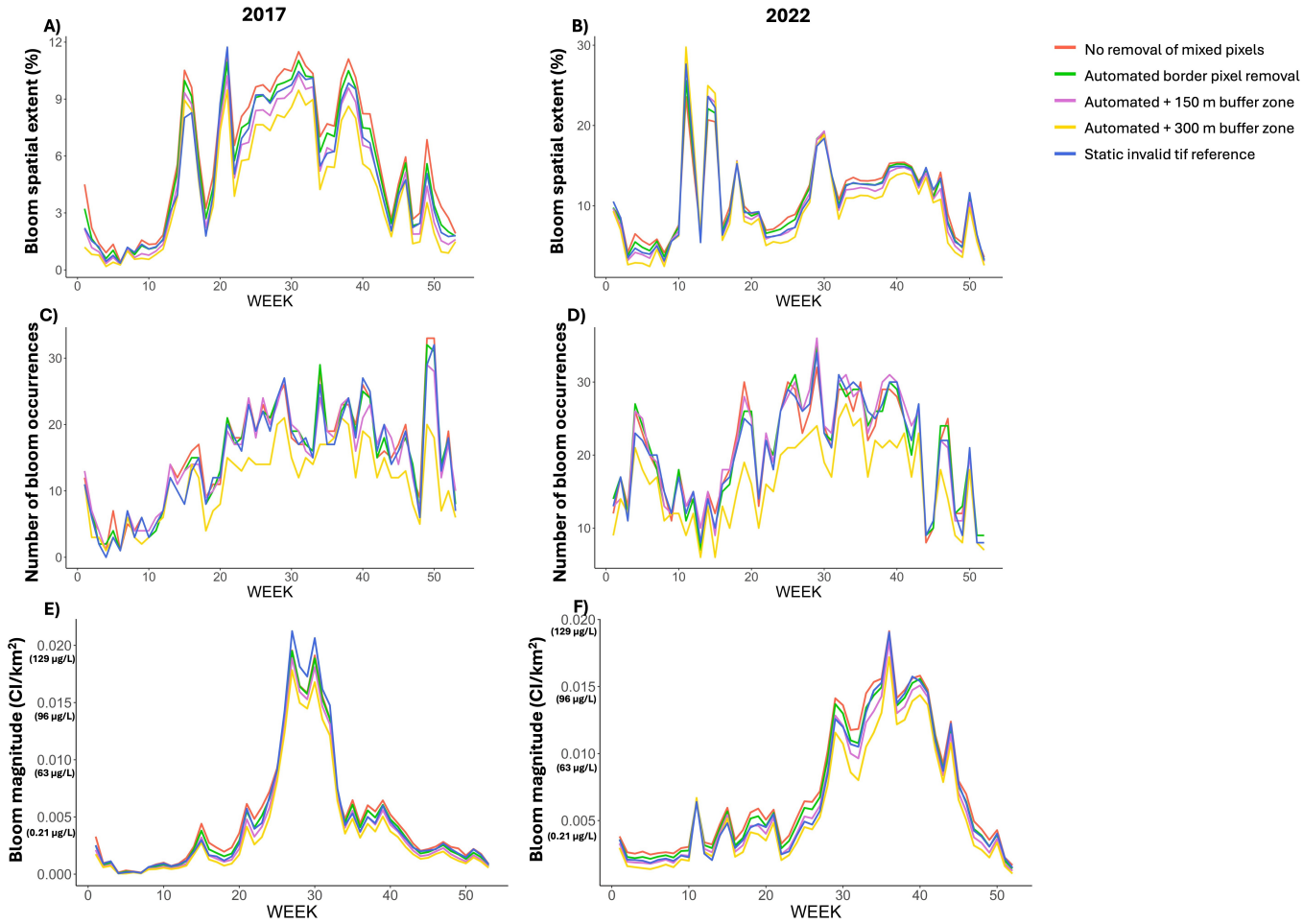

Figure 4.6: Comparison of weekly bloom spatial extent (% of total pixels) for A) 2017 and B) 2022, weekly number of bloom detection occurrences (defined by DN > 0, > 10% pixels) for C) 2017 and D) 2022, and weekly bloom magnitude for E) 2017 and F) 2022 using different mixed pixel removal methods for 83 lakes.

In summary, the sets of data processed for this comparison included:

1. Retention of both mixed and ice pixels
2. Removal of ice pixels only

3. Removal of mixed pixels only using a 150 m buffer zone of the polygon border
4. Removal of ice pixels and mixed pixels only using a 150 m buffer zone of the polygon border

4.4.2 Impact of snow and ice flagging on resulting bloom metrics

We compared data processed with or without snow and ice pixel removal. Meanwhile, mixed pixels (Section 4.3) were either not removed or were removed by automated masking with a 150 m buffer zone (Figure 4.7).

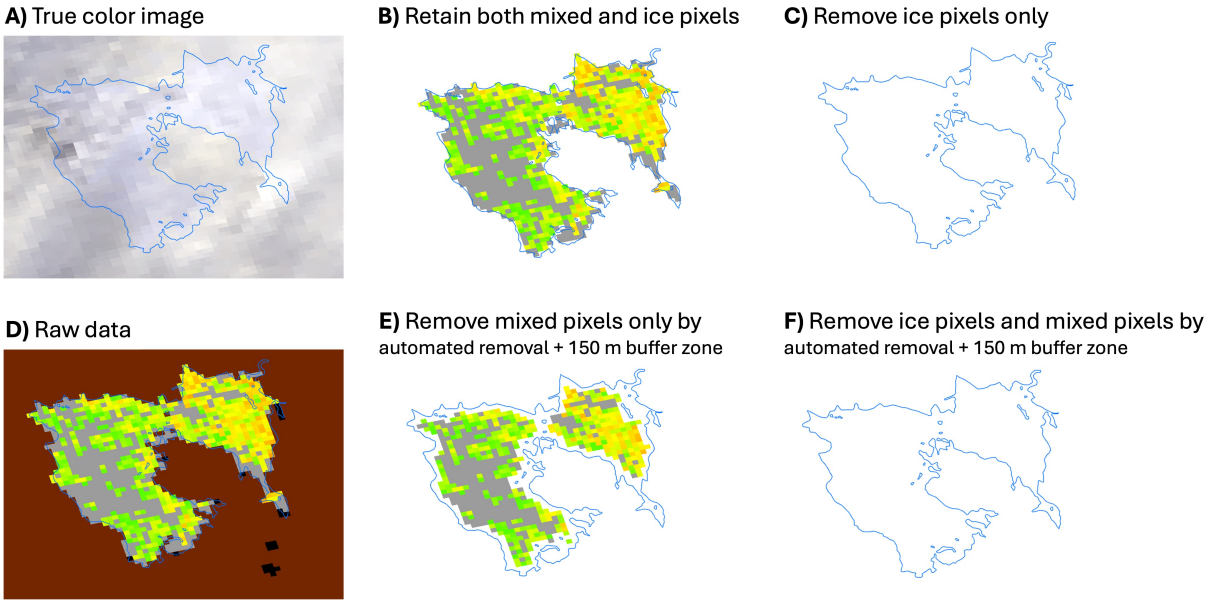

Figure 4.7: Different processing methods for invalid pixels retention/removal using Clear Lake Reservoir located in the Modoc National Forest in week 1 of 2020 as an example. A) True color image of the reservoir indicating mostly ice and snow coverage. B) No removal of border (i.e., potential mixed land-water pixels) and ice pixels. C) Removal of ice pixels only. No resolvable pixel remains after removing ice pixels. D) Raw data from CyAN. Brown pixels in the raw data indicate land pixels flagged by CyAN project data processing procedures (DN = 254) and black pixels indicate 'No data' pixels flagged by the CyAN project (DN = 255). E) Removal of potential mixed land-water pixels by use of a 150 m buffer zone. F) Removal of both potential mixed land-water pixels and ice pixels results in no resolvable pixels remaining after all ice pixels were removed.

The removal of ice pixels expectedly reduced the number of resolvable pixels each year (non-striped bars vs. striped bars in Figure 4.8A). The annual total number of bloom occurrences was also reduced as with average annual bloom detection frequency for most years (non-striped bars vs. striped bars in Figure 4.8B, C), but the average annual bloom alert frequency was

mostly similar between removal/retention of ice pixels, and can, at times, be higher with the removal of ice pixels (non-striped bars vs. striped bars for 2019 and 2022 in Figure 4.8D). The higher number of bloom occurrences and annual bloom frequencies suggested that ice pixels sometimes had DN > 0 values. However, these bloom occurrences and cyanobacterial detections were likely false positives since optical signals from ice and snow can sometimes be wrongly interpreted as cyanobacteria by the CIcyano algorithm (Urquhart and Schaeffer 2020). These false positive pixels were mostly of lower DN (DN < 132) and thus not included in the bloom alert category, which led to mostly similar annual bloom alert frequency between removing or retaining snow and ice pixels. The occasional increase in bloom alert frequency from the removal of snow and ice pixels is likely due to a decrease in total number of valid pixels (through the removal of snow and ice pixels) while the number of pixels defined as bloom alert (DN $\geq$ 132) remained the same. 2019 was a particularly interesting year as both annual bloom detection frequency and annual bloom alert frequency increased with snow and ice pixel removal. The higher annual bloom frequency after the removal of ice pixels were due to weeks 11 and 12 (3/17/2019 – 3/30/2019) specifically, when the removal of ice pixels led to a dramatic increase in bloom frequency. Comparing results from mixed pixel removal by automated masking and 150 m buffer zone with or without ice pixel removal, the removal of snow and ice pixels in those 2 weeks led to a ~60% decrease in pixels identified as 'no bloom' (DN = 0), but the number of pixels identified as 'bloom' (DN > 0) remained similar (~8,000 – 9,000). The dramatic increase in bloom frequencies for these 2 weeks subsequently led to an overall increase in annual bloom frequency after ice pixel removal in 2019. Nevertheless, overall comparison of the removal/retention of ice pixels indicate that imagery data collected in the presence of ice/snow, as determined by referencing data from NSIDC, commonly had DN > 0, and the retention of these pixels leads to a higher annual number of bloom occurrence and bloom frequency.

Investigation of the weekly statewide data for bloom spatial extent, number of occurrences, and bloom magnitude indicate that the largest differences in ice pixel removal vs. retention tended to occur in weeks 0 – ~20, and weeks ~45 – 52 (Figure 4.9), which corresponds to the colder weeks of the year when snow and ice are present. The removal of ice pixels consistently led to lower numbers of bloom occurrence in these weeks (Solid line below dashed lines Figure 4.9C, D). Notably, the results presented here show that the removal of ice pixels can lead to either higher or lower bloom extent. Bloom spatial extent was generally lower or similar after ice pixel removal for the first and last weeks of the year, while bloom spatial extent was higher after ice pixel removal ~weeks 10 – 20 and around week 45 for both years 2017 and 2022 (Figure 4.9A, B, solid line above dashed lines ~weeks 10 – 20 and around week 45). Lower or similar bloom extent after ice pixel removal is likely due to a decrease in number of pixels identified as bloom (i.e., DN > 0) and total number of resolvable pixels, while higher bloom extent is likely due to similar number of bloom pixels (i.e., none or little of the removed ice

pixels were bloom pixels), but a decrease in the total number of resolvable pixels after removing snow and ice pixels (Figure 4.10). This pattern of lower or similar bloom extent at the colder time of the year (weeks 0 – 10 and weeks 45 – 52) indicated that most of the snow and ice pixels removed during those periods had DN > 0, while the tendency of higher bloom extent after snow and ice pixel removal occurring ~weeks 10 – 20 and around week 45 indicated that the snow and ice pixels during those periods had DN = 0. Therefore, as expected, snow and ice pixels had the highest risks of leading to false-positive bloom detection at the coldest times of the year. This is reflected by the larger difference in number of lakes with blooms during weeks 0 – ~10 and weeks ~45 – 52 (Figure 4.9C and D). In contrast, bloom magnitude was generally higher after ice pixel removal in this study. As bloom magnitude is based on both the summation of bloom intensity (as indicated by CIcyano value after conversion from DN) as well as the area of waterbody (number of resolvable pixels multiplied by pixel area), higher bloom magnitude after ice pixel removal is likely due to the decrease in waterbody area after removal of ice pixels while the summation of CIcyano values remain similar.

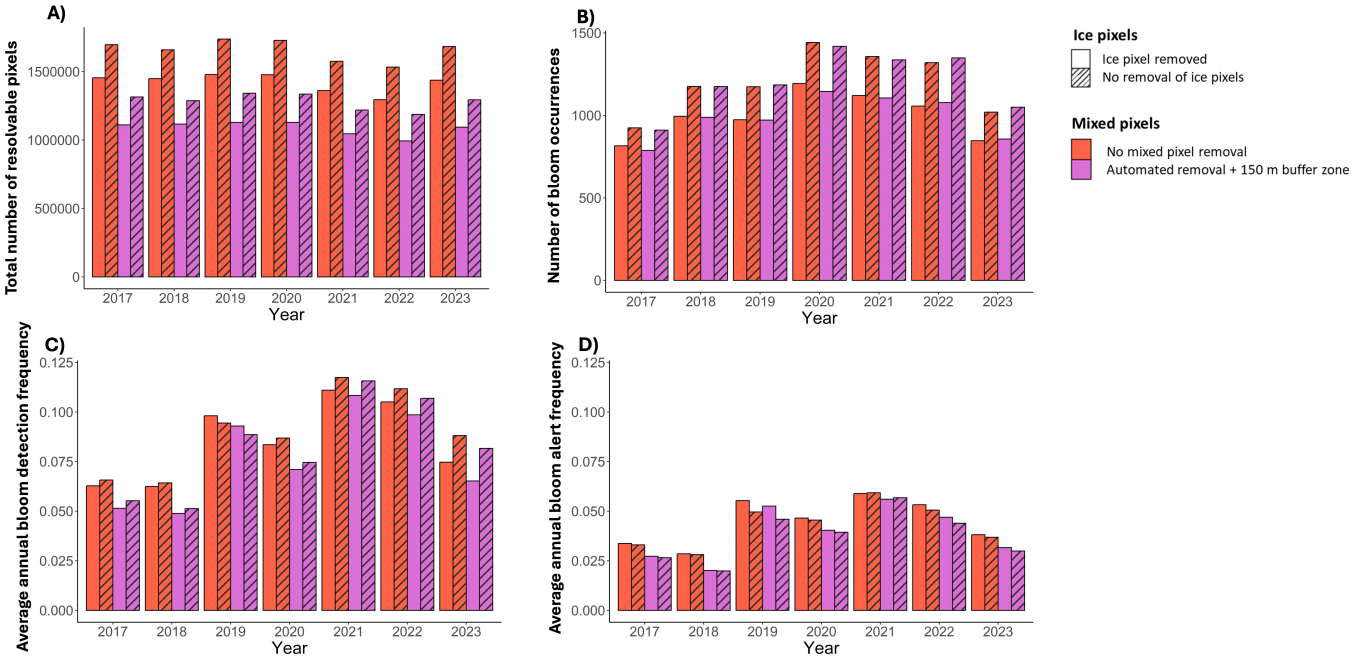

Figure 4.8: Comparison of different combinations of mixed pixel and ice inclusion or exclusion methods (as indicated by bar color and pattern combinations on A) total number of pixels processed in the year; B) annual number of weekly bloom detection occurrences (DN > 0 and spatial extent > 10%); C) average annual bloom detection frequency (DN > 0); and D) average annual bloom alert frequency (DN ≥ 132) for 83 waterbodies. Frequency is a proportion.

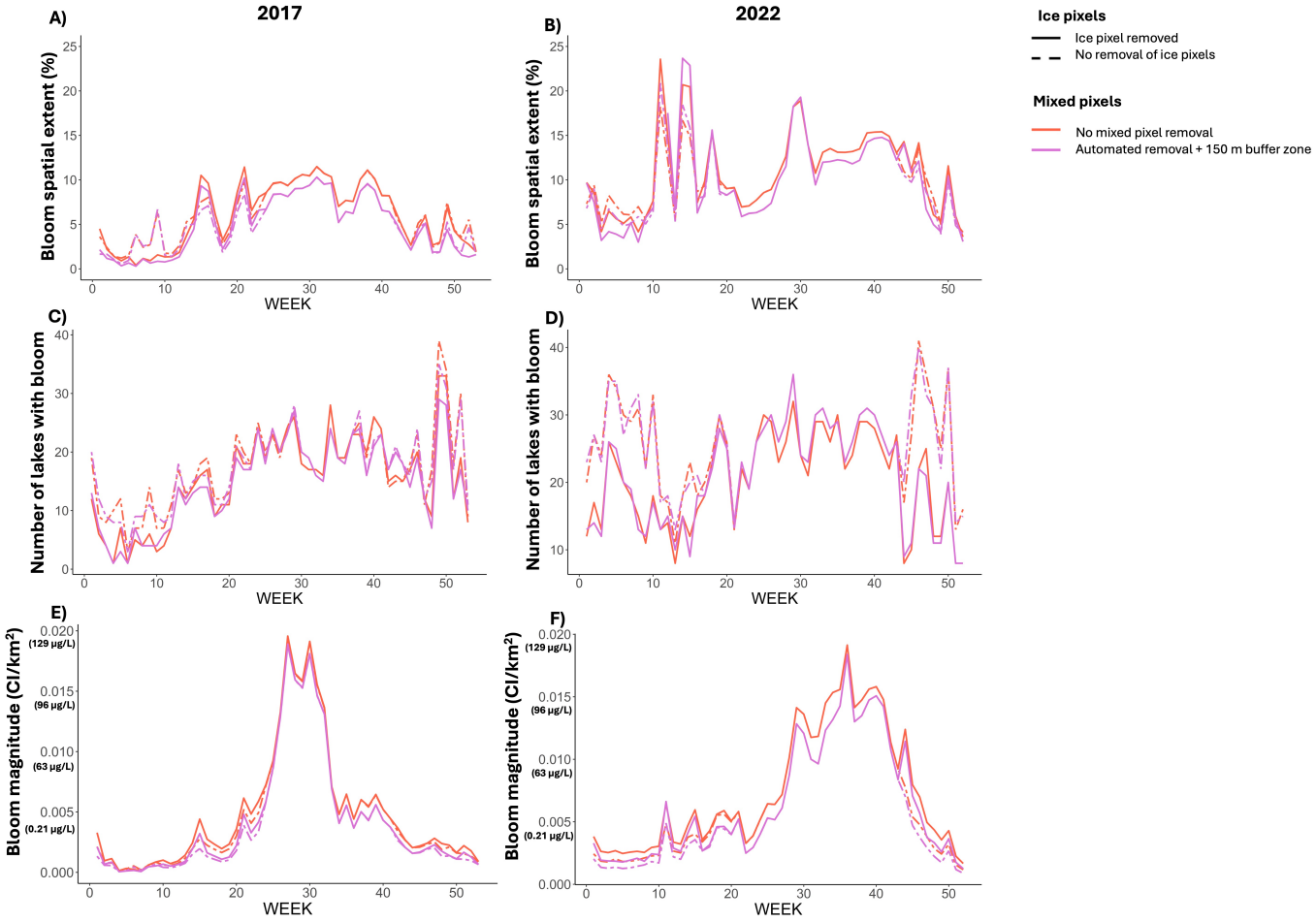

Figure 4.9: Comparison of weekly bloom detection spatial extent (% of total pixels) for A) 2017 and B) 2022, number of lakes with bloom detection occurrences (defined by DN > 0, > 10% pixels) for C) 2017 and D) 2022, and weekly bloom magnitude for E) 2017 and F) 2022 using different combinations of mixed pixel and ice pixel removal methods for 83 lakes.

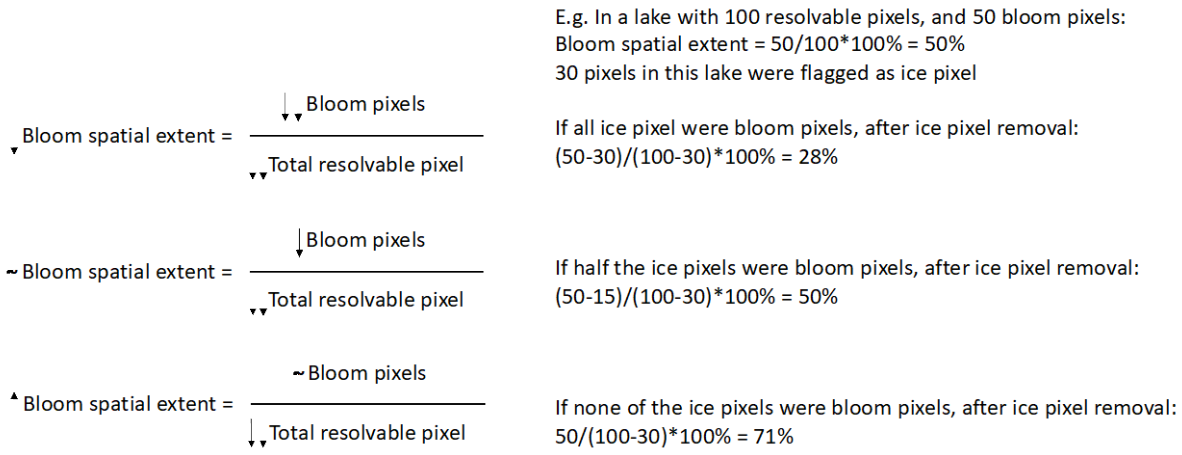

Figure 4.10: Depiction and example of how bloom spatial extent can be affected by the removal of ice pixels. Down arrow indicates a decrease, '~' indicates similar result, and up arrow indicates an increase.

A summary comparison of the percentage difference between the retention/removal of mixed and/or ice pixels indicated how results can be affected by seasons. Summer, for example, appeared to be least affected by the different choices for invalid pixels as it tended to have the smallest range and mean for choosing to retain mixed and/or ice pixels compared to removing both types of invalid pixels (Table 4.3). The pervasiveness of negative minimum values in the table, which indicates lower metric values when mixed and/or ice pixels are retained, showed that the removal of both types of invalid pixels can lead to higher bloom metric values despite the generally lower values in overall comparisons (Figure 4.5, Figure 4.6, Figure 4.8, Figure 4.9).

Table 4.3: Summary of the range and mean (in parentheses) percentage difference (difference between two methods/average x 100%) in number of bloom occurrence (DN > 0 and spatial extent > 10%), number of bloom alert occurrences (DN ≥ 132 and spatial extent > 10%), bloom spatial extent (DN > 0), bloom alert spatial extent (DN ≥ 132), and area normalized magnitude between different methods compared to removing both mixed and ice pixels from 2017 – 2023. Higher values indicate larger differences of the method compared to removal of both mixed and ice pixels. Negative values indicate lower metric value of method compared to removing both mixed and ice pixels. Winter = Weeks 1 – 9 and Weeks > 48; Spring = Weeks 10 – 22; Summer = Weeks 23 – 35; Fall = Weeks 36 – 48. Mixed pixels were removed via automated masking of a 150 m buffer zone from the edge of the polygon. Ice pixels were removed via reference from 4 km ice/snow coverage data from NSIDC.

Metric	Method	Overall	Winter (Dec – Feb)	Spring (Mar – May)	Summer (Jun – Aug)	Fall (Sep – Nov)
Occurrence	Retain both mixed and ice pixels	-33 – 110% (19%)	-10 – 110% (46%)	-19 – 100% (24%)	-24 – 19% (-3%)	-33 – 105% (9%)
Occurrence	Remove mixed pixels only	-5 – 105% (19%)	5 – 95% (44%)	-5 – 71% (20%)	-5 – 10% (1%)	-5 – 105% (12%)
Occurrence	Remove ice pixels only	-47 – 42% (1%)	-21 – 42% (4%)	-26 – 42% (4%)	-26 – 21% (-3%)	-47 – 42% (-2%)
Bloom alert occurrence	Retain both mixed and ice pixels	-27 – 108% (18%)	-14 – 108% (28%)	-27 – 108% (17%)	-27 – 68% (13%)	-14 – 95% (12%)
Bloom alert occurrence	Remove mixed pixels only	-15 – 58% (3%)	0 – 58% (8%)	-15 – 44% (4%)	0 – 0% (%)	-15 – 58% (2%)
Bloom alert occurrence	Remove ice pixels only	-28 – 111% (14%)	-14 – 111% (19%)	-28 – 97% (12%)	-28 – 69% (14%)	-14 – 55% (10%)
Extent	Retain both mixed	-340 – 111%	-90 – 92% (24%)	-340 – 111%	-87 – 25% (5%)	-39 – 56% (14%)

Metric	Method	Overall	Winter (Dec – Feb)	Spring (Mar – May)	Summer (Jun – Aug)	Fall (Sep – Nov)
	and ice pixels	(11%)		(2%)		
Extent	Remove mixed pixels only	-274 – 135% (2%)	-80 – 75% (12%)	-274 – 135% (-4%)	-16 – 1% (0%)	-46 – 44% (0%)
Extent	Remove ice pixels only	-170 – 44% (10%)	-35 – 44% (15%)	-170 – 27% (5%)	-88 – 25% (5%)	-8 – 29% (13%)
Bloom alert extent	Retain both mixed and ice pixels	-265 – 62% (7%)	-91 – 62% (5%)	-265 – 43% (5%)	-127 – 38% (11%)	-94 – 36% (7%)
Bloom alert extent	Remove mixed pixels only	-217 – 57% (-5%)	-99 – 57% (-4%)	-217 – 33% (-8%)	-17 – 0% (0%)	-108 – 25% (-6%)
Bloom alert extent	Remove ice pixels only	-123 – 37% (12%)	-33 – 31% (11%)	-122 – 32% (13%)	-123 – 37% (11%)	-20 – 27% (12%)
Magnitude	Retain both mixed and ice pixels	-139 – 72% (8%)	-52 – 72% (5%)	-139 – 51% (6%)	-31 – 42% (15%)	-105 – 31% (7%)
Magnitude	Remove mixed pixels only	-118 – 68% (-4%)	-62 – 68% (-3%)	-118 – 45% (-6%)	-16 – 0% (0%)	-116 – 25% (-6%)
Magnitude	Remove ice pixels only	-61 – 41% (13%)	-9 – 22% (10%)	-61 – 29% (13%)	-31 – 41% (15%)	-20 – 30% (13%)

Retaining snow and ice pixels led to higher differences in the number of bloom occurrences in all seasons except summer (See first 2 rows for 'occurrence' in Table 4.3). The maximum percentage differences were high (> 70%), as well as the mean difference particularly for winter and spring (> 20%). These relatively high positive values indicated that the retention of ice pixels resulted in higher and likely erroneous counts of bloom occurrence. This is most likely due to false detection based on known optical interferences (Urquhart and Schaeffer 2020). It is possible, however, that some of these instances may reflect true cold water bloom detections as there are field based observations of blooms around and under ice on record (Graham et al., 2008, Reinl et al., 2023). Paired satellite and field observations under snow and ice conditions could help to untangle this dynamic, however it is generally accepted that the OLCI sensor is unlikely to be able detect these particular events due to the interference of the ice spectral signature (Urquhart and Schaeffer 2020). The retention of mixed pixels (See last row for 'Occurrence' in Table 4.3), in comparison, had relatively stable differences throughout the

seasons. The percentage differences observed for the number of bloom alert occurrences were similar to that of bloom occurrences, with the retention of ice pixels leading to higher counts of bloom alert occurrences in winter, spring, and fall, but the increase in number is of a lower magnitude compared to bloom occurrences, implying that only a portion of the removed ice pixels have DN ≥ 132. Waterbodies with higher number of bloom and bloom alert occurrences from retaining ice pixels were mostly waterbodies located at higher latitudes where snow and ice cover are more prevalent (Figure 4.11).

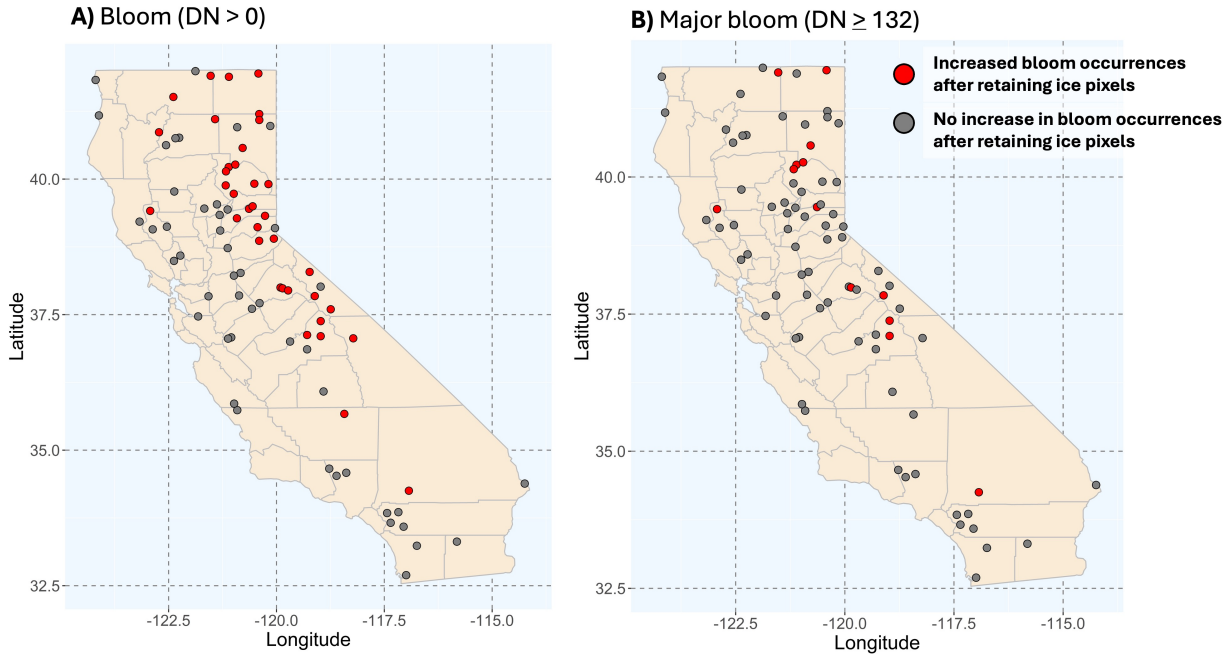

Figure 4.11: Waterbodies with higher number of A) bloom detection occurrences (DN > 0) and B) bloom alert occurrences (DN ≥ 132) from ice pixel retention during winters of 2017 – 2023 (Weeks 1 – 9 and Weeks > 48).

While the retention of mixed and/or ice pixels had a wide range of percentage difference in bloom detection and alert spatial extent (e.g., range of -340 – 111% for overall extent from retaining both mixed and ice pixels, first row and third cell of 'Extent' in Table 4.3), the mean percentage differences were mostly around or lower than 10%. This suggests that the extreme range of percentage differences observed for bloom detection and alert extent may be limited to a small number of weeks. The retention of snow and ice pixels led to higher differences in winter and spring for both bloom detection and alert spatial extent (See second row for both 'Extent' and 'High extent' in Table 4.3), with a generally negative mean value. Such negative mean values indicate that the removal of snow and ice pixels in winter led to higher bloom extent values, which occurs when snow and ice pixels were mostly non-bloom pixels (DN = 0; Figure 4.10). An exception to the generally negative percentage difference was bloom extent in

Winter (12%). This positive mean percentage indicates that the retention of snow and ice pixels in winter led to higher bloom extent values while removing snow and ice pixels led to lower bloom extent values, which occurs when snow and ice pixels were mostly bloom pixels (DN > 0; Figure 4.10). This again demonstrated the strong effect of snow and ice pixels leading to false-positive indication of blooms during the winter season, but such effect is only limited to winter and bloom extent evaluation. The retention of snow and ice pixels led to a mean negative percentage difference for bloom alert extent even for winter, indicating that ice pixels overall did not have DN $\geq$ 132.

Interestingly, the retention of mixed pixels also led to higher mean percentage difference in bloom extent in winter and fall compared to spring and summer, indicating that mixed pixels were more often identified as bloom pixels (DN > 0) in the winter and fall which resulted in higher bloom extent values during these seasons. These mixed pixel 'blooms' could be accumulation of cyanobacteria from the previous summer's blooms nearer to the shoreline, or alternatively, could represent interference from land vegetation. Field observations would help to identify the likely cause of these differences.

The removal of ice pixels led to a general increase in bloom magnitude in all seasons, while the removal of mixed pixels led to a general decrease in bloom magnitude in all seasons, as indicated by the mostly negative mean percentage difference from ice pixel retention (second row of 'Magnitude' in Table 4.3) and the consistent positive mean percentage difference from mixed pixel retention (third row of 'Magnitude' in Table 4.3). This suggests that the removed ice pixels did not have a strong contribution to the bloom intensity of the waterbody (i.e., sum of CIcyano for all resolvable pixels), but the removed mixed pixel did have a strong contribution to the bloom intensity (i.e., removed mixed pixels had high CIcyano values).

4.5 Temporal Compositing of the Data

4.5.1 Types of temporal composites

Daily and 7-day maximum (7D max) value imagery composites can be downloaded from the CyAN project data portal. SFEI also serves 10-day maximum (10D max) value composites from NOAA for download through the FHAB Satellite Analysis Tool (https://fhab-api.sfei.org/). Note that NOAA processes the satellite data with minor variations compared to the CyAN project, so CIcyano values from the two sources may have slight differences.

Custom 7D max, 7D median (7D med), and 10D max composites were generated from daily CyAN data products from 2017 - 2022 to compare the use of different types of temporal composites. 7D max composites were generated from daily imagery rather than using pre-generated 7D max composites from CyAN to ensure that the custom generated composites

were comparable in their processing. The range of dates for the 7D composites were from 1/1/2017 – 12/31/2022, while the range of dates for the 10D composites were from 1/1/2017 – 12/30/2022. The 7D 90th percentile was also tested for a smaller subset of the data and yielded nearly identical results to the 7D max (data not shown).

These satellite image composites were processed using the custom California waterbodies shapefile (with 238 waterbodies), automated mixed pixel masking with 150 m buffer zone, and ice-pixel removal using 4 km ice data from NSIDC. Waterbodies included in the analysis was further restricted to 'Suitability for trend analysis'= 'Suitable' and 'Water Type' = 'Lake + Reservoir' (Refer to section 4.2 for details), which resulted in a final count of 158 lakes and reservoirs. The use of the 150 m buffer zone resulted in the loss of 37 smallest and/or more narrow waterbodies from the inventory as they had 0 resolvable pixels in the 6 years of data analyzed when the additional buffer zone was applied.

4.5.2 Impact of temporal aggregation on bloom metrics

Comparison between 7D and 10D max composites indicated that 7D and 10D composites yielded similar annual bloom detection frequencies despite a significant reduction in number of resolvable pixels analyzed by 10D composites (Figure 4.12). While the annual bloom detection and alert frequencies for 10D max composites are consistently higher than the 7D max composites (Figure 4.12C, D), it should be noted that the scale for the frequencies (y-axis) were small, and so the overall difference between the two types of composites are small. This is more apparent when comparing the average annual frequencies for no bloom (DN = 0), in which the y-axis is from 0 – 1 (Figure 4.12E). The total number of bloom occurrence each year was lower for 10D max composites, which is due to the smaller number of composites analyzed each year (i.e. 52 – 53 per year for 7D composites; 36 – 37 per year for 10D composites).

While the number of resolvable pixels analyzed between 7D max and 7D med composites were the same, 7D med composites consistently led to more conservative values of various bloom metric values. Number of bloom occurrences, annual bloom frequencies, and bloom spatial extent were all consistently lower from 7D med composites than 7D max composites (Figure 4.12, Figure 4.13). Despite quantitative differences, 7D max and 7D med generally followed the same interannual patterns, with the exception of a few bloom events particularly in 2022 that were characterized in the 7D max dataset and not in the 7D med dataset (Figure 4.13D, presence of red 7D max peaks and the lack of 7D med peaks). However, without large-scale frequent field observations, it is difficult to understand which dataset most accurately characterized bloom dynamics during those events.

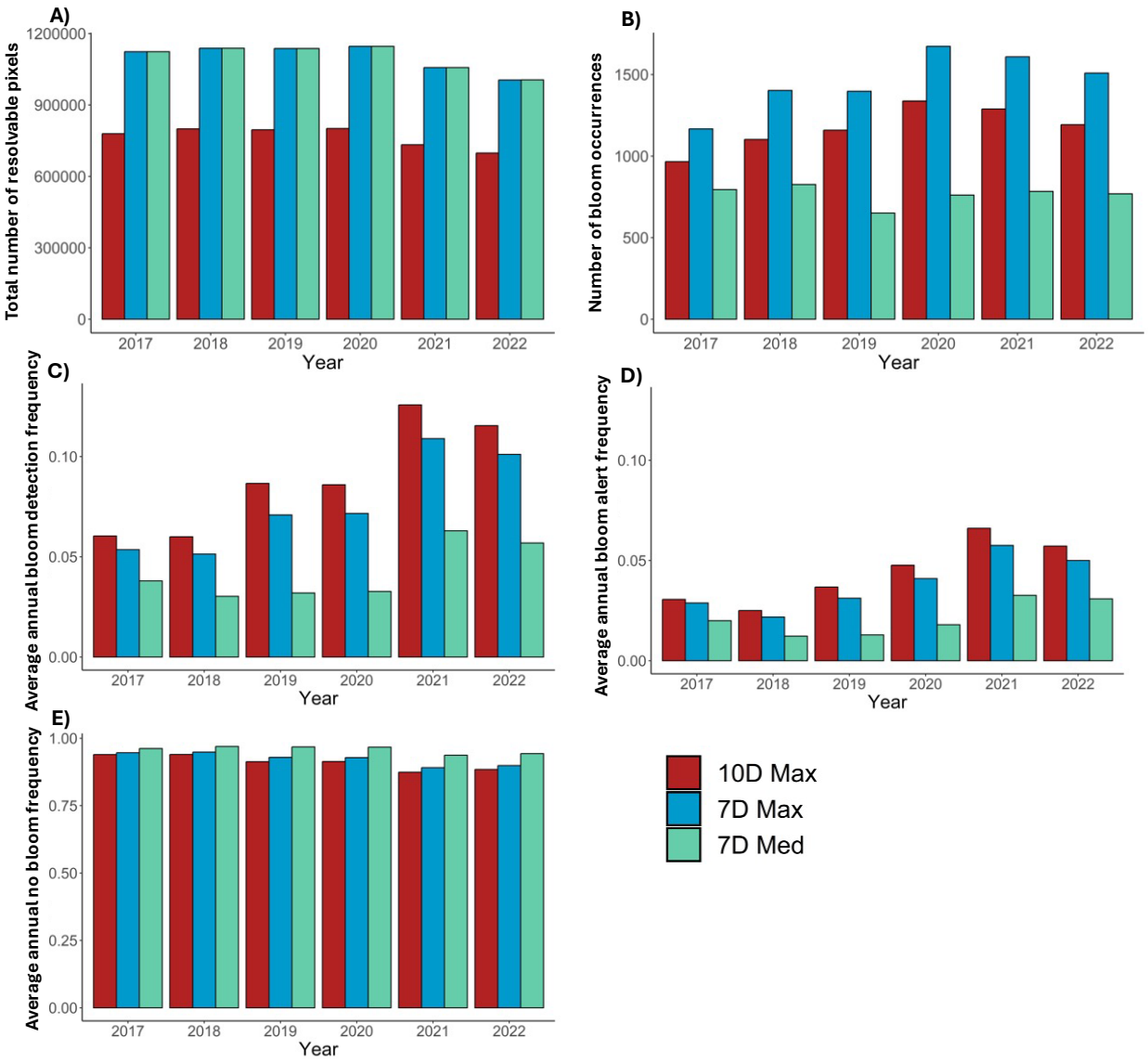

Figure 4.12: Comparison temporal compositing approaches (as indicated by bar color) on A) total number of pixels processed in the year; B) total number of bloom occurrence (DN > 0 and spatial extent > 10%); C) average annual bloom frequency (DN > 0); D) average annual frequency of bloom alerts (DN ≥ 100); and E) average annual frequency of no bloom (DN = 0) for 83 waterbodies. Bloom frequency is presented as a proportion. Mixed pixels were removed by automation with 150 m buffer zone and ice pixels were removed using 4 km ice data from NSIDC as reference.

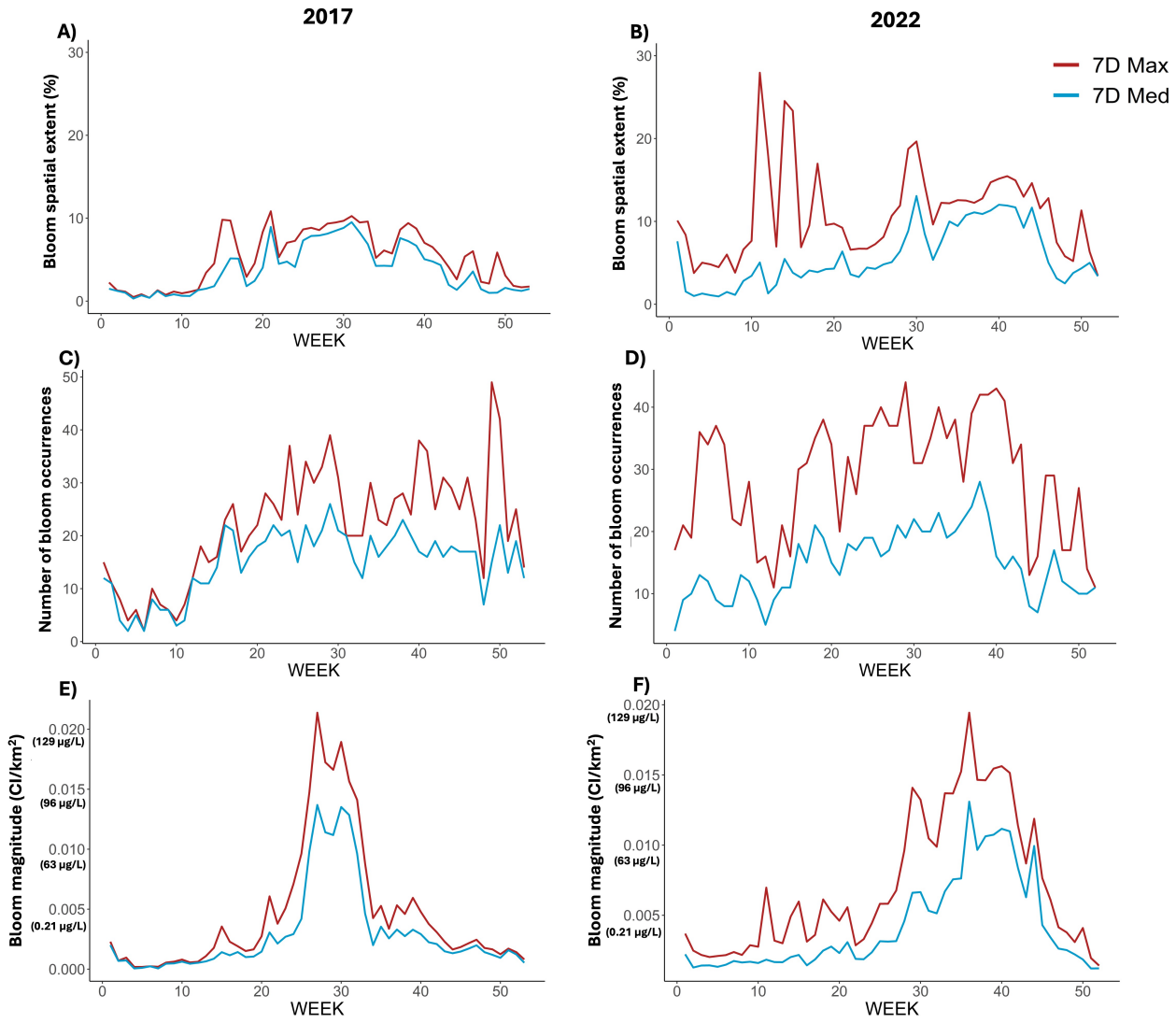

Figure 4.13: Comparison of weekly bloom spatial extent (% of total pixels) for A) 2017 and B) 2022 and number of bloom occurrences (defined by DN > 0, > 10% pixels) for C) 2017 and D) 2022 between 7D max and 7D med composites for 83 waterbodies.

4.6 Conclusions and recommended procedures

The previous sections investigated the removal of mixed land-water and ice pixels, illustrating how different processing methods can affect the results and bloom metric values generated from the same satellite imagery data, and that the retention of invalid pixels may lead to false positives of bloom occurrence (e.g., ice pixels as shown in Figure 4.8).

Based on these results, *it is recommended to remove ice pixels before any further analyses*. We found evidence of false positives for cyanobacteria detection (e.g., DN >0), albeit at relatively low levels. Furthermore, the increased detection of blooms in locations known to have ice and

snow suggests these are most likely erroneous results. These findings are congruent with previously published work (Urquhart and Schaeffer 2020).

Table 4.4: A summary of recommended standardized procedures for satellite imagery processing to assess FHABs.

Processing Step	Recommendations for Standard Procedure
Selection of resolvable waterbodies	Use California custom shapefile using attributes 'Suitability for trend analysis'= 'Suitable' and 'Water Type' = 'Lake + Reservoir'
Removal of potential mixed land-water pixels	Automated removal of pixels along waterbody polygon boundary
Removal of potential snow and ice pixels	Removal of potential ice pixels using 4 km ice reference map
Temporal compositing of imagery data	Use 7-day maximum temporal composites

The decision of how to process mixed land-water pixels ultimately depends on the stakeholder goals, and considerations on the balance between data confidence and the retention of potentially mixed or land influenced pixels (which may or may not provide realistic data) must be made based on these goals. The variation in the actual water boundary for waterbodies due to climatic/weather conditions (e.g., drought or heavy rain) further complicates this decision. *For the status and trend assessment goals of the FHAB Program, the boundary pixel removal approach is recommended* since it strikes an appropriate balance between inclusion of as many waterbodies as reasonable and removal of edge pixels that are very likely mixed pixels. This approach strikes a balance between conservative mixed pixels removal and retaining enough pixels for robust further analysis on a larger number of waterbodies. This approach of automated masking is also scalable for future efforts involving high spatial resolution imagery. The use of the more conservative 150 m and 300 m boundary removal procedure with the custom California waterbodies shapefile resulted in the loss of 70 smaller/more narrow waterbodies from the inventory as they had 0 resolvable pixels in the 6 years of data analyzed. Automated removal of boundary pixels is an efficient method to remove mixed pixels, but an important caveat is that it is highly dependent on the accuracy of the polygons in the shapefile. This balance seems appropriate given that testing against conservative mixed pixel removal processes (i.e., the use of 150 and 300 m buffer zones) yielded generally similar results. Importantly, the retention of nearshore pixels may allow for more detection of cyanobacterial bloom occurrence by including pixels closer to the shoreline and may thus serve well for detecting events that may not extend into open water zones, particularly when paired with field-based investigations or special studies. For efforts such as event response where a field-based assessment would be conducted as a follow up to the satellite observations, retention of

the more nearshore pixels might be desirable to capture these more nearshore data. When resources are available to implement recommendations, future efforts should consider comparing buffer zone data from *in-situ* observations to satellite observations to further understand the comparability (or not) of these two data types.

A more conservative approach to the removal of possible mixed pixels via a buffer zone may be most suitable for applications where stringent data certainty is required. Data obtained from the CyAN project will include land pixel flagging (DN = 254), but it is unknown how well the algorithm performs with mixed land-water pixels, and whether there is a threshold of the proportion of land vs. water for a pixel to be flagged as land. It is also unknown how the inclusion of small amounts of land within the overall pixel area influences the data quality, since the measurement is integrating over a relatively large (300 x 300 m) area. Making a static TIF for reference can allow for more confidence in the inclusion of only pixels that contain 100% water but requires a large investment of time and labor and is not feasible to repeat routinely. Even with manual curation, a static reference would only reflect a snapshot of the water boundary condition at the timepoint that the reference images were collected. Thus, this approach is only recommended when a high level of customization or effort is focused over a specific time period. The use of a buffer zone allowed a more conservative approach for removing mixed pixels that is less intensive to generate but may lead to undesired removal of otherwise resolvable waterbodies with just a few pixels. Nevertheless, a buffer zone of 150 m often led to similar results as using a static, manually curated reference (Figure 4.5, Figure 4.6) despite the likely difference in some of the actual pixels removed between the two methods (i.e., the pixels removed by the static reference is different than those removed by the automated masking with 150 m buffer zone). This approach may strike a good balance between conservative mixed pixel removal and retaining enough pixels for proper further analysis given that the waterbodies are big/wide enough to allow such a buffer around its border. Otherwise, as observed in section 4.5.1, smaller and/or more narrow waterbodies will have no resolvable pixels after mixed pixel masking with a 150 m buffer zone.

Different types of temporal composites (e.g., 7D vs. 10D or max vs. med) can also lead to different values of bloom metrics. Temporal composites made from maximum and median values led to similar or greater differences in bloom metric timeseries than variations in 7D and 10D compositing, with composites using median values giving more conservative results. Therefore, the decision to use maximum vs. median once again depends on end user goals, and whether a more sensitive (e.g., bloom incident early warning) or conservative (e.g., for Water Board management decisions) approach is preferred. Ibelings et al. (2021) had suggested that in cases where toxin data are not available, the maximum cyanobacteria biomass could aid in the estimation of maximum expectable cyanotoxin concentrations, thus *the recommendation for the FHAB Program is to use maximums in temporal composites, particularly for management*

applications. Trend analyses or comparisons across different time periods, however, should be more cautious in temporal aggregation approach due to differences in observation frequency before and after the launch of Sentinel-3B (see Section 5.2). Results from 7D and 10D composites were comparable, and the decision to use 7D vs. 10D may just be user preference and data availability (if composites were already available). *For the FHABs Program, 7D composites are recommended since these translate readily into weeks, which are readily interpretable when consolidated into the downstream data metrics such as bloom occurrence.*

5. KEY GRAPHICS AND LINKAGE TO MANAGEMENT QUESTIONS

This section provides example key graphics from further analysis and a summary of processed OLCI imagery data, applying the data metrics described in Chapter 3, and their linkages to management questions described in Chapter 2. The focus of the analyses in this chapter is to address the status and trends questions described in Table 2.1, with a particular emphasis on statewide and waterbody specific status, patterns, and trends. Analyses in this chapter focused on cyanobacterial blooms and thus CIcyano and cyanobacterial Chl-a were applied as indicators, although a similar analysis could also be conducted using total Chl-a should the end user wish to define blooms more broadly. 7D max data from 2017 – 2023 were processed using the custom California waterbodies shapefile with 238 waterbodies, automated masking (no buffer zone), and ice pixel removal using the 4 km ice data from NSIDC, according to the data processing recommendations in Chapter 5. The waterbodies included in the analysis were restricted to 'Suitability for trend analysis'= 'Suitable' and 'Water Type' = 'Lake + Reservoir' (Refer to section 5.2 for details), which resulted in a final count of 158 waterbodies. Both blooms detection (defined as DN > 0) and bloom alerts (defined as blooms associated with $\geq$12 µg/L cyanobacterial Chl-a, DN $\geq$ 132) were examined. Status was assessed comparatively to the running average of each metric in relation to the observations of the specific year (i.e., 2023). Formal trend calculations were conducted in section 5.2.

5.1 Status and patterns analysis

5.1.1 Statewide

Cyanobacterial bloom conditions in 2023 were reduced compared to 2021 – 2022 as the overall bloom detection and alert occurrences and frequency were below the average of the mean for 2017 – 2023 (Figure 5.1). Satellite imagery data indicated that the number of bloom detections and alerts started to increase in 2018 and were highest in 2020 (for bloom detections) and 2021 (for bloom alerts; Figure 5.1A, Figure 5.1B). Annual bloom detection and alert frequency started to increase in 2019, and peaked in 2021 (Figure 5.1C, Figure 5.1D). Changes from 2017 and 2018, however, should be interpreted with caution given that observational frequency across this time period increased with the addition of the Sentinel-3B satellite in 2018 (Coffer et al., *submitted*). Due to these changes, observed increases may be due to increased temporal coverage of both Sentinel-3A and 3B satellites rather than true environmental change (see section 5.2 for additional discussion).

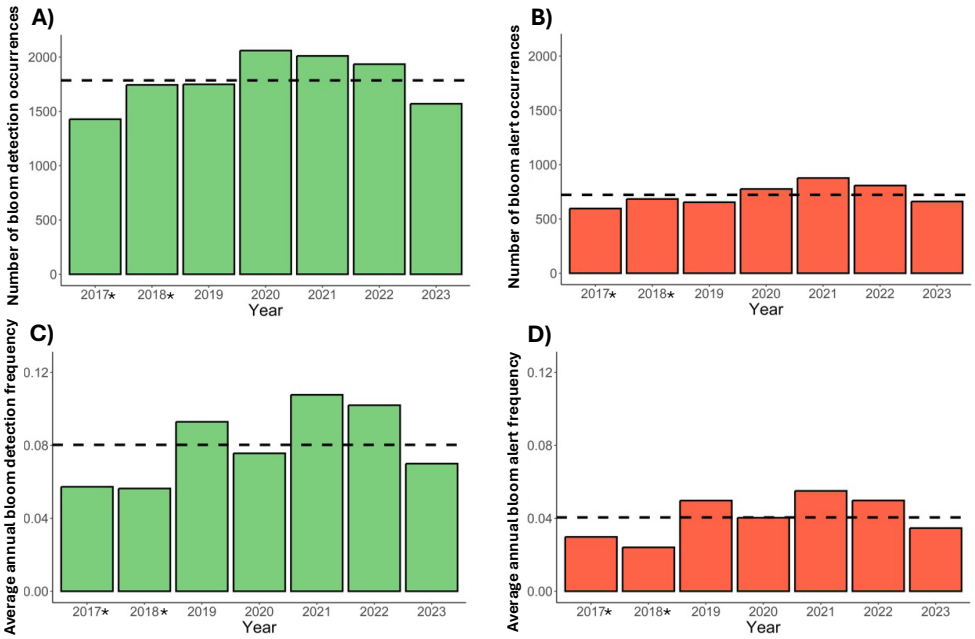

Figure 5.1: Comparison of A) total number of bloom detection occurrences per year, B) total number of bloom alert occurrences per year, C) average annual bloom detection frequency, and D) average annual bloom alert frequency for lakes and reservoirs suitable for trend analysis (Total = 158). Frequency is a proportion. Dashed line indicates the 7-year average. The * symbol indicates years during which satellite observational frequency was lower compared to the rest of the timeseries.

Seasonal analysis indicated that summer and fall are usually the seasons with the highest bloom spatial extent, number of lakes with blooms, and bloom magnitude, while winter usually has the lowest bloom extent, magnitude, and number of occurrences. Average bloom detection and alert extent was particularly high for summers of 2019 and 2021 (red bars in Figure 5.2A, B), which fit with the increased bloom detection and alert annual frequencies observed for those years (Figure 5.1C, D). Bloom magnitude for summer of 2017 and fall of 2021 and 2022 were particularly high even though bloom detection and alert extent was not particularly high for those seasons. This may indicate high intensity blooms were present (as suggested by the magnitude metric) for a small number of pixels (thus constraining bloom alert extent to modest values) during those seasons. The above average number of bloom occurrences observed for years 2020, 2021, and 2022 (Figure 5.1A) were due to increased occurrences of winter and spring blooms (blue and green bars in Figure 5.2C), highlighting that blooms can occur year-round in many waterbodies in California.

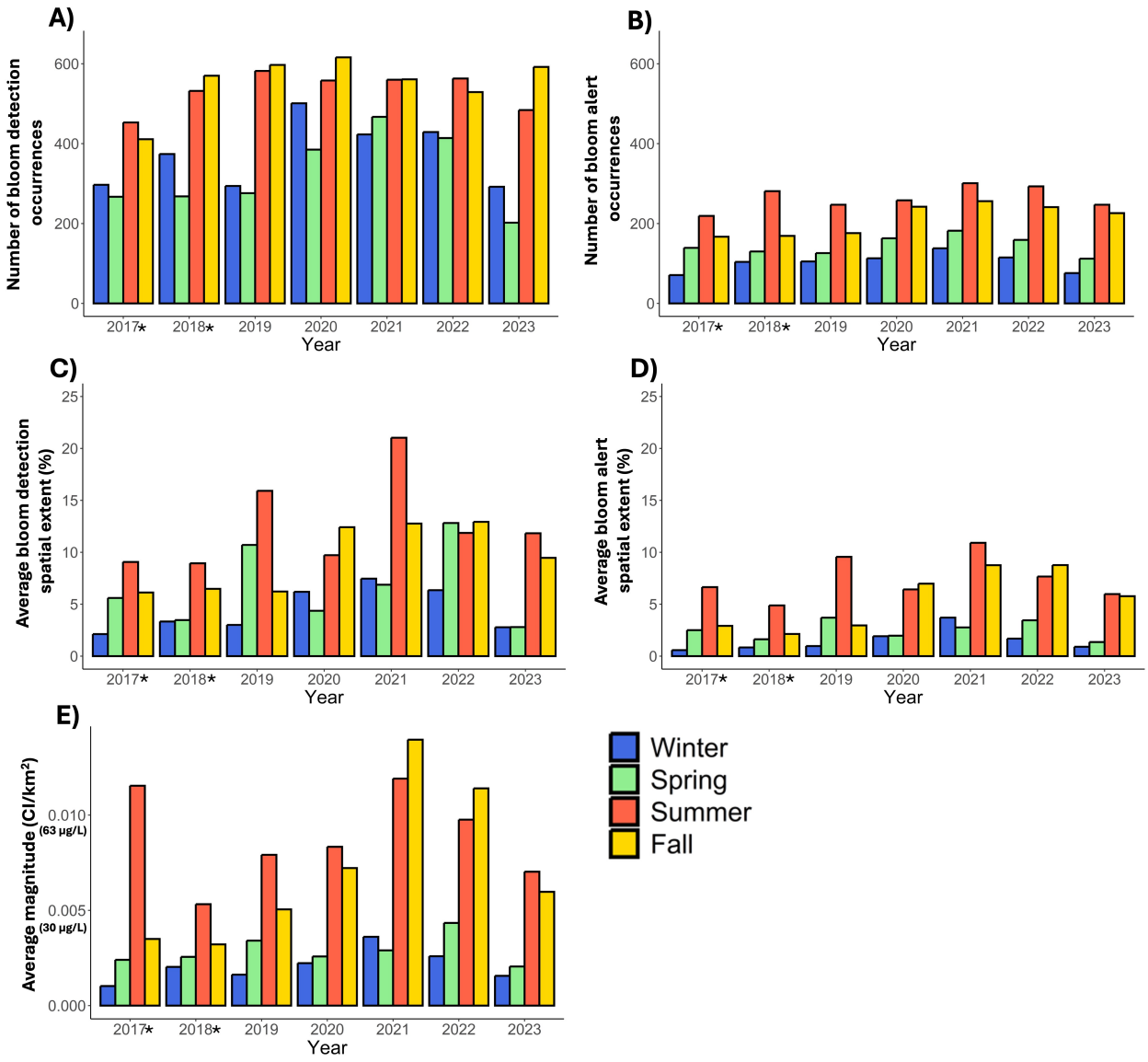

Figure 5.2: Seasonal average of A) bloom detection and B) bloom alert spatial extent. The count of of C) bloom detection and D) bloom alert occurrences, and E) bloom magnitude for lakes and reservoirs suitable for trend analysis (Total = 158). Winter = Weeks 1 – 9 and Weeks > 48; Spring = Weeks 10 – 22; Summer = Weeks 23 – 35; Fall = Weeks 36 – 48. The * symbol indicates years during which satellite observational frequency was lower compared to the rest of the timeseries.

Weekly analysis of the state-wide bloom detection and alert spatial extent indicated that spatial extent generally remained stable, peaking around ~8 – 10% for bloom detections, and ~5 – 10% for bloom alerts (Figure 5.3C, D). Spikes of increased bloom detections and alerts (> 20% for bloom detections and >15% for bloom alerts) were observed for a small number of weeks in 2019 (weeks 11 – 12, 32 – 33) and 2021 (weeks 29 – 32), which indicated there was an overall increase in bloom alert spatial extent throughout the state during those periods. Weekly number of bloom occurrences and bloom magnitude (Figure 5.3A, B, E) followed the seasonal

pattern of higher numbers in the middle of the year (summer and fall) compared to the start and end of the year.

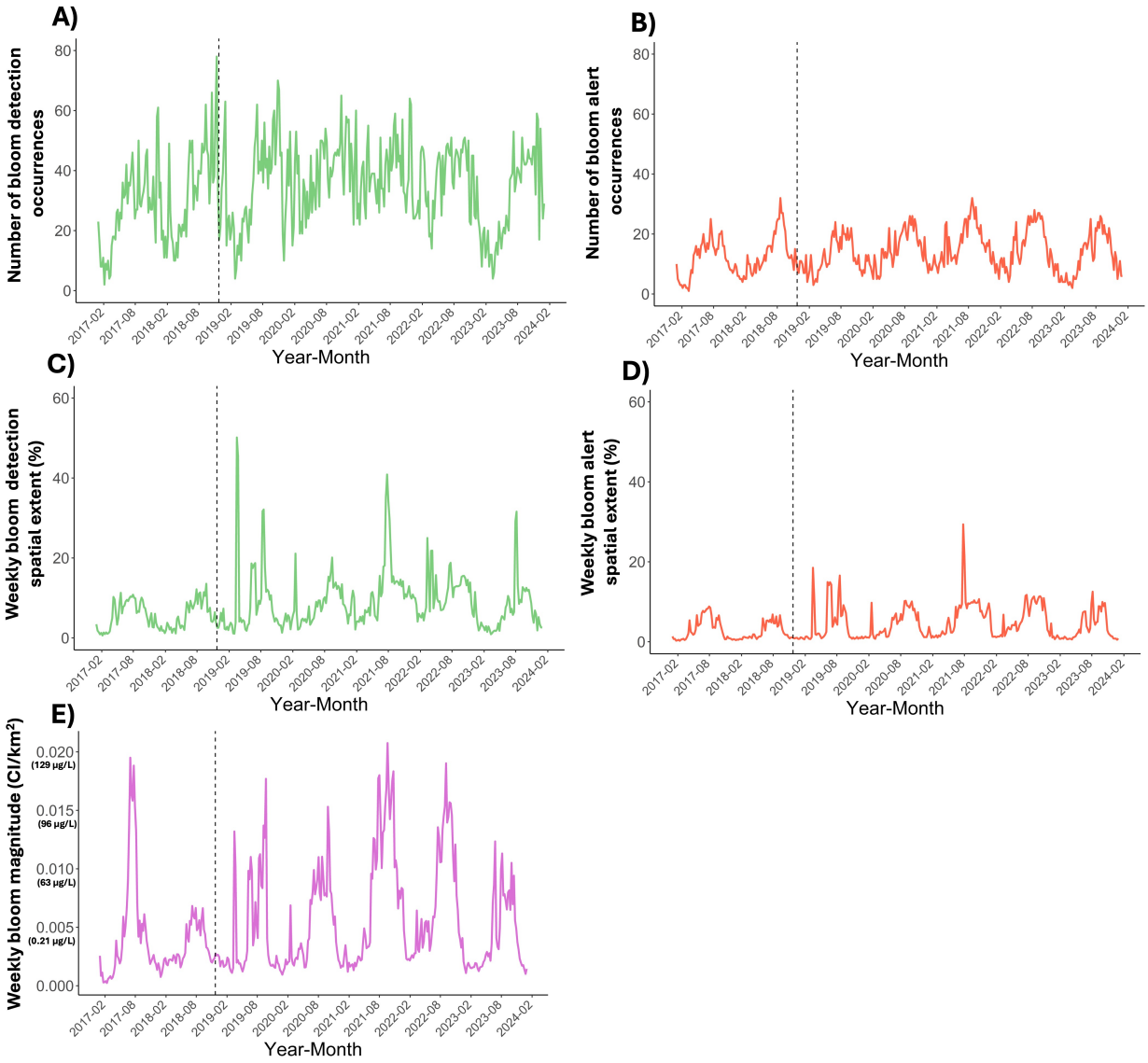

Figure 5.3: Weekly number of lakes and reservoirs suitable for trend analysis (Total = 158) with A) bloom detections and B) bloom alerts. Weekly C) bloom detection spatial extent and D) bloom alert extent, and E) weekly bloom magnitude. Winter = Weeks 1 – 9 and Weeks > 48; Spring = Weeks 10 – 22; Summer = Weeks 23 – 35; Fall = Weeks 36 – 48. Black vertical dashed line indicates Sentinel-3B reaching its nominal orbit phased 140° from Sentinel-3A and essentially doubling the observational frequency of the Sentinel-3 satellite series.

Bloom metrics were used to identify if any geographic patterns in FHABs were apparent in resolvable lakes across California. There were no visually apparent geographic patterns or

geographic clusters of lakes with similar numbers of bloom detection occurrences in 2023 (Figure 5.4). Instead, waterbodies bloom occurrence counts for each resolvable waterbody on the map were randomly distributed across the state when considering both detectable and alert level blooms. Despite the overall lower number of bloom occurrences in 2023 compared to previous years, there were a number of waterbodies that had higher number of bloom detection and alert occurrences in 2023 compared to the 7-year average for that waterbody (Figure 5.5). These waterbodies with increased number of cyanobacterial bloom occurrences were again present throughout the state.

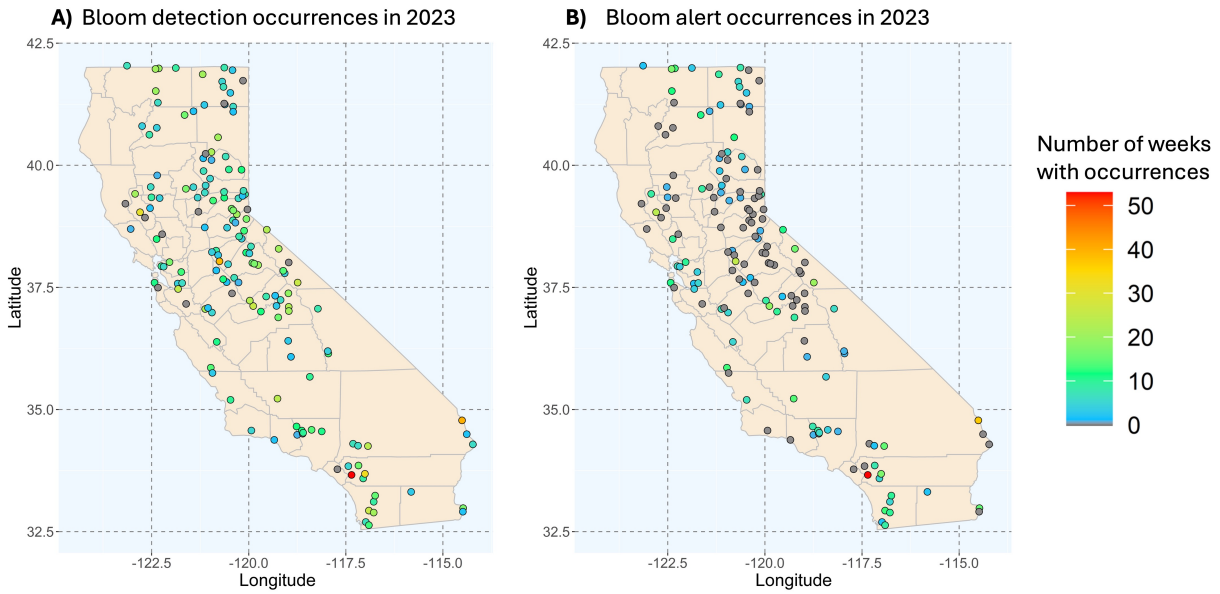

Figure 5.4: Location and counts of A) bloom detection (DN > 0) and B) bloom alert (DN ≥ 132) occurrences for lakes and reservoirs suitable for trend analysis (Total = 158) in 2023. The count of weekly bloom detection and alert occurrences are indicated for each resolvable lake with warmer colors indicating a higher number and cooler colors indicating a lower number.

All lakes and reservoirs suitable for trend analysis (n = 158) were sorted into 6 distinct categories according to their total number of bloom occurrences each year and the number of lakes in each category were compared across years (Figure 5.6). The categories used were: 1) 0 weeks with a bloom; 2) 1-4 weeks with a bloom; 3) 5-12 weeks with a bloom; 4) 13-24 weeks with a bloom; 5) 25-35 weeks with a bloom; and 6) >36 weeks with a bloom. The category '5 – 12 weeks' consistently had the highest number of waterbodies throughout the 7 years analyzed (green bars in Figure 5.6), while the category '> 36 weeks' consistently had the lowest number of waterbodies (red bars in Figure 5.6). This indicated that there were only a small number of waterbodies that had bloom occurrences > 70% of the year. A decrease in waterbodies in the categories '25 – 36 weeks' and '> 36 weeks' for 2023 indicated a reduction in the number of

waterbodies that had persistent blooms throughout the year and implied some improvements in the statewide bloom situation. Notably, 2023 is the year following record levels of rainfall in late 2022 and throughout early 2023 which may have influenced bloom dynamics.

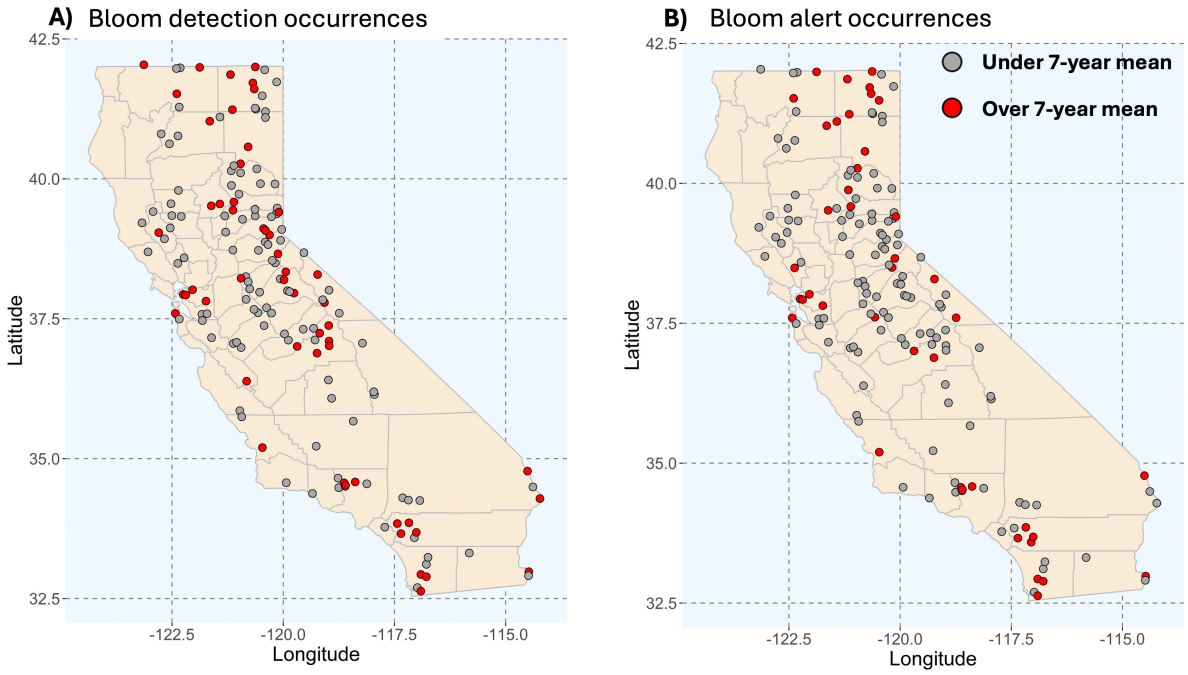

Figure 5.5: Locations of lakes and reservoirs suitable for trend analysis (Total = 158) with the number of A) bloom detection occurrences and B) bloom alert occurrences in 2023 above their each respective 7-year (2017 - 2023) mean number of bloom detection and alert occurrences.

5.1.2 Individual lakes

This section demonstrates the application of similar status and patterns analyses on individual waterbodies using Clear Lake as an example.

An average of 28 near-surface, open water bloom occurrences were detected at Clear Lake between the years of 2017 – 2023 (Figure 5.7) Most of the detected blooms (>90%) were classified as major (high) blooms with DN $\geq$ 132 constituting over 10% of the resolvable lake pixels. However, bloom pixels were generally not homogenous in the lake, and limited to certain regions. For example, the median DN values for the pixels in the upper arm of Clear Lake was 0 in 2023 (Figure 5.8 left panel), indicating there were no cyanobacterial bloom detection more than half of the year when the pixels were resolvable. This type of spatial pattern was observed in most years (data not shown), and matches *in situ* observations at the lake (Smith et al., 2023). Looking at the maximum pixel DN values, on the other hand, showed that blooms were detected across the entire lake for at least some portion of the year, as indicated by each

pixel experiencing an elevated DN value (DN $\geq$ 132) for at least 1 week in 2023 (Figure 5.8 right panel).

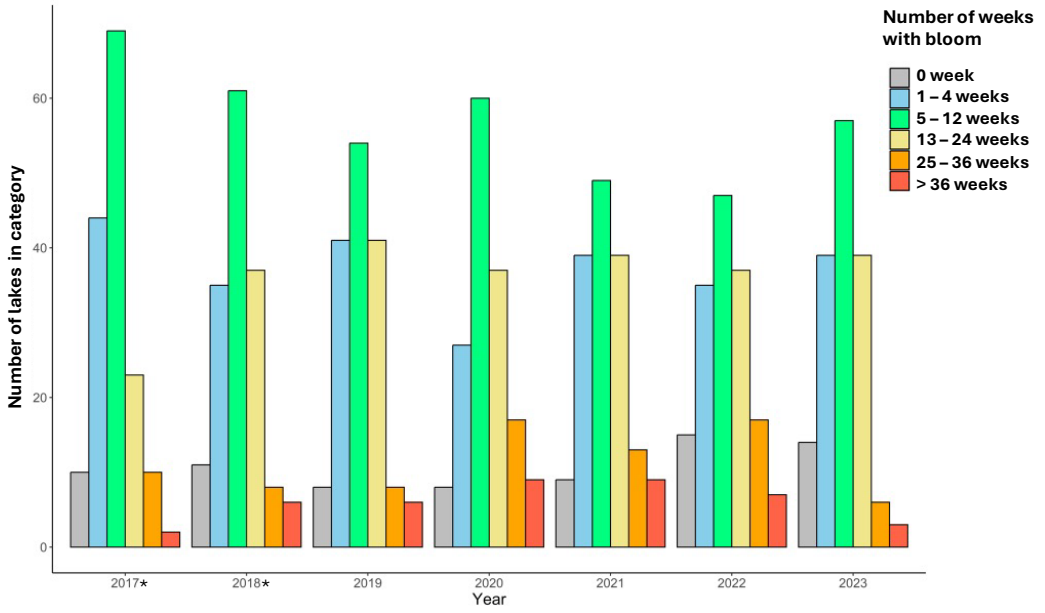

Figure 5.6: Number of lakes and reservoirs suitable for trend analysis (Total = 158) sorted into categories indicating number of weeks with bloom occurrences from 2017 – 2023. The * symbol indicates years during which satellite observational frequency was lower compared to the rest of the timeseries.

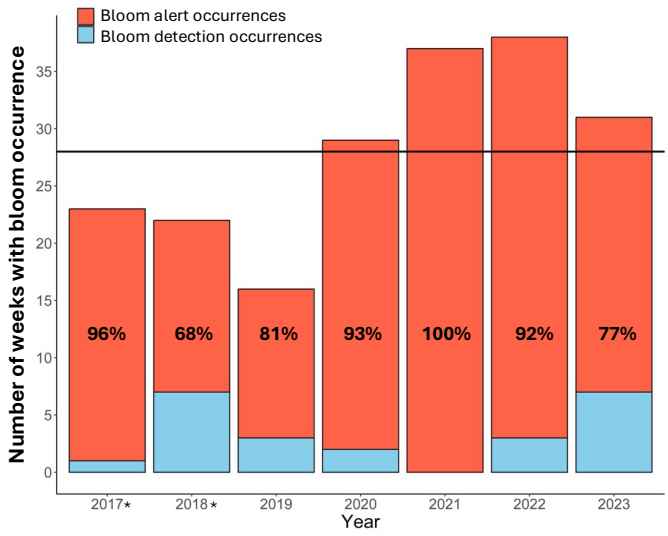

Figure 5.7: Annual number of bloom alerts (red bars) and detections (blue bars) in Clear Lake from 2017 - 2023. Black horizontal line indicates the 7-year mean number of bloom occurrences. Percentages indicate the percentage of bloom alert occurrences among total bloom occurrences. The * symbol indicates years during which satellite observational frequency was lower compared to the rest of the timeseries.

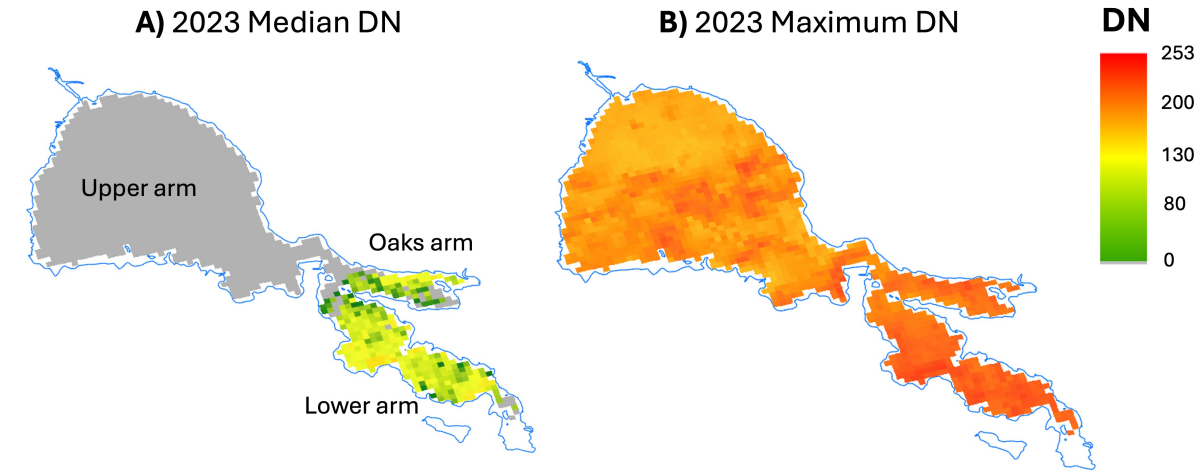

Figure 5.8: Median and maximum pixel DN values for all resolvable weeks for each pixel in Clear Lake over the year in 2023.

Weekly bloom spatial extent values indicated similar annual patterns and values over the 7-year period (Figure 5.9A, B), in which bloom extent was low (<5%), a rapid increase to ~100% around May, and then another rapid decrease to <5% around September/October. This suggested that blooms generally occurred and covered large areas of the lake relatively rapidly, with similar rapid bloom termination. Sharp peaks in bloom magnitude, on the other hand, suggests that high bloom intensity periods were generally shorter in duration in comparison (Figure 5.9C).

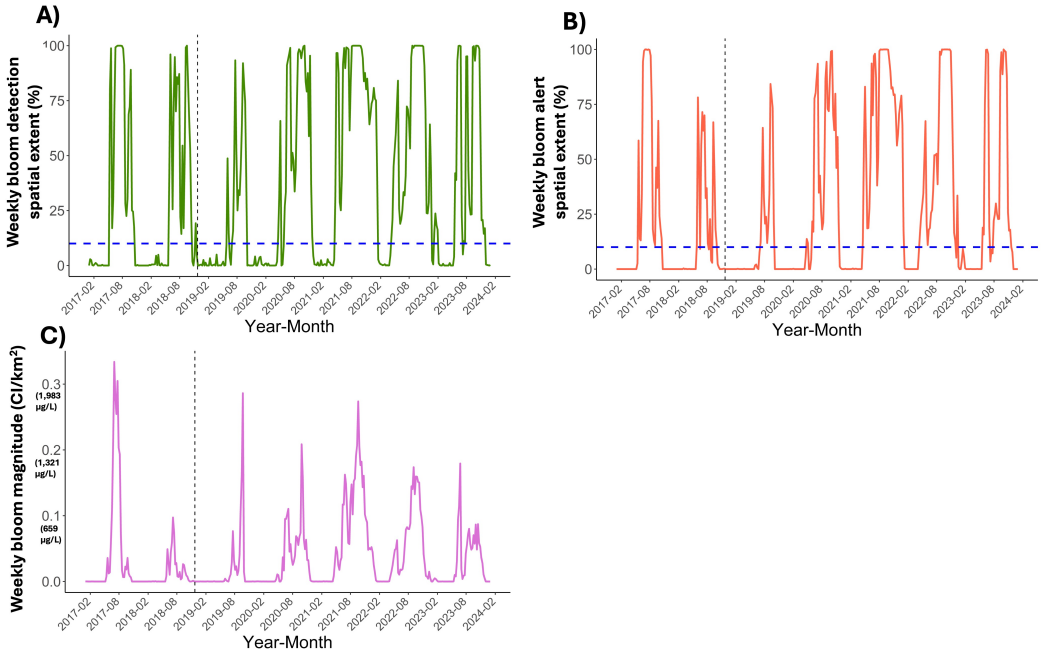

Figure 5.9: A) Weekly bloom detection extent, B) bloom alert extent, and C) bloom magnitude at Clear Lake. Blue horizontal dashed line indicates 10% bloom occurrence spatial threshold i.e., values above the horizontal dashed line indicates a bloom occurrence. Black vertical dashed line indicate a doubling of the observational frequency of the Sentinel-3 satellite series.

5.2 Statistical Trend Analyses

5.2.1 Statistical approaches for calculating trends and baselines

Annual and monthly trends in bloom metrics can be investigated using a variety of statistical approaches. One of the most used approaches for trend analysis of water quality data, including bloom metric data, is the Mann-Kendall approach (Hirsch et al., 1982; Urquhart et al., 2017). This non-parametric approach tests for monotonic trends in metrics. It handles missing data well (as long as missing data is not in systematic intervals), making it ideal for imagery data where clouds, ice, glint and other factors can result in sporadic missing observations. This is a more appropriate approach for monotonic trends than linear regression analysis which is more suited for normal data distributions and can be more strongly influenced by missing data and serial dependence (Hirsch et al., 1982). Both trend analysis approaches are sensitive to autocorrelation.

The Mann-Kendall test can detect trends by calculating differences in signs between earlier and later time points and determines if there is a consistent increase or decrease in sign values over the time series for each individual month or across years. It has two different variations, the Seasonal Mann-Kendall Test and the Mann-Kendall Test, which are suitable for timeseries that have seasonality and for annualized timeseries, respectively. Trends for monthly or weekly time series that encompass a seasonal cycle are best estimated using a Seasonal Mann-Kendall test to account for seasonality across months (Hirsch et al., 1982). The Mann-Kendall test follows the same principles but is best suited for annual data inputs. The Kendall's tau value, which ranges between -1 and 1, indicates the relative strength of the trend. Values closer to one indicate a stronger trend and values closer to -1 indicate a weaker trend. A value of 0 indicates no trend. The sign of the Kendall's tau value indicates the directionality of the trend with a positive value indicating an increasing trend and a negative value indicating a decreasing trend. Trends are considered significant at p-values ≤ 0.05.

A waterbody can exhibit no directional trend for several reasons, including because it has frequent or persistent blooms or because blooms are exceptionally rare. Determining baseline conditions can allow the addition of supplemental context to waterbodies without significant directional trends and delineate these two scenarios (i.e., frequent blooms or rare blooms). Baseline conditions can be determined by calculating the historical average condition or other measure of central tendency for a given bloom metric for either the individual waterbody or the larger population of waterbodies of the same type. Individual waterbodies with no trend for a given bloom metric can then be compared to the historical average condition for the larger population to determine if the historical average of that metric for that waterbody is below or

above baseline condition. These waterbodies can then be characterized as being above or below the historical baseline for that metric to better contextualize no trend data.

5.2.2 Temporal compositing considerations for trend analyses with imagery data

Trend assessments that include Sentinel-3 observations from 2017 and 2018 should be interpreted with caution given that observational frequency across this time period varies. Sentinel-3A was launched by the ESA in February 2016, and Sentinel-3B was launched in April 2018 and reached its nominal orbit on 23 November 2018 (Clerc et al., 2020). Thus, during mid-2018, observation frequency changed from 1-2 images per week (2016-2018) to near-daily imagery beginning in November 2018 due to the added satellite platform (Schaeffer et al., 2022). Recent work by Coffer et al. (*submitted*) demonstrated that maximum 7D composites can lead to observed increases in trends due to increased observational frequency through the addition of satellites to the constellation rather than true environmental change since the increased observational frequency increases the opportunity to observe extreme values. Two different approaches exist to mitigate this effect. First, using a measure of central tendency for weekly compositing, such as using a median or mean instead of a maximum, helps to reduce this effect when using a timeseries that includes periods where the satellite constellation changes (e.g., going from one platform to two). Second, limiting trend analyses to a period where the number of satellite platforms is consistent (e.g., only the period when 2 satellite platforms are in use) can also reduce this effect. Notably, 7D maximum composites are a routinely available product, while 7D medians need to be generated by the user, thus the choice between these options should balance these added data processing considerations.

Trends in both annual and seasonal bloom metrics were tested for the period of 2019-2023 when two satellite platforms were consistently in use. This time period represents five years of year-round observations, meeting the minimum time period of five years recommended by Hirsch (1988) for water quality trend analyses. 158 resolvable lakes from the custom California waterbody inventory were used based on using the analysis inclusion criteria of water type = “Lake” and Use type = “All”. These waterbodies were processed using the procedures recommended in Chapter 5 and used automasking (i.e., no buffer zone) for edge pixel removal and 4km ice removal. The Mann-Kendall test was applied to the annualized bloom metrics, annual bloom frequency and annual number of weeks with bloom occurrences. The seasonal Mann-Kendall test was applied to bloom metrics that contained seasonality within the timeseries, categorized as monthly average bloom extent and number of bloom occurrence per month. Trends were considered significant at p-values ≤ 0.05. Lakes with non-significant trends (p-value > 0.05) were evaluated according to the statewide average for that bloom metric from

2019-2023 and were categorized as either no change and above that baseline or no change and below that baseline.

Overall, significant directional trends for annual bloom frequency and annual number of bloom occurrences were rare, with only 1 (<1% of lakes assessed) lake demonstrating a significant decreasing trend in annual bloom frequency. More directional trends were observed for monthly average bloom extent and monthly number of weeks with bloom occurrences. A total of 10 (6%) and 7 (4%) lakes had decreasing trends for bloom extent and weeks with bloom occurrences, respectively. Less lakes had an increasing trend (Table 5.1). For all bloom metrics, most lakes had no statistically significant trend. Depending on the bloom metric, between 45-63 lakes (28% - 40%) were above the statewide baseline for a given metric, while 90-106 (57% - 67%) were below the baseline.

Table 5.1: Summary of results of trend analysis for the period of 2019-2023 using 7D maximum temporal composites. The number of lakes in each trend category are summarized in the table.

Bloom Metric Tested	Increasing Trend (*n*)	Decreasing Trend (*n*)	No change, lake above baseline (*n*)	No change, lake below baseline (*n*)	Statewide baseline
Annual Bloom Frequency	0	1	51	106	0.15
Annual number of weeks with bloom occurrence	0	0	63	95	11.80
Monthly Average Bloom Extent	4	10	45	99	14.45
Number of weeks with bloom occurrences per month	3	7	58	90	0.98

6. SATELLITE REMOTE SENSING DATA QUALITY FOR ASSESSMENT OF FHABS

The purpose of this chapter is to define the elements that comprise a data quality assessment and to review existing literature on these elements for OLCI imagery data. Guidance for data quality assessments of environmental data was developed by US EPA and SWAMP integrated it into their quality assurance manual (EPA, 2000b). The focal elements of this literature review conducted for this report were selected based on the SWAMP quality assurance manual since the FHAB Program intends to develop a quality assurance project plan governing the use of Satellite Remote Sensing data for water quality management decisions using the information summarized in this report.

6.1 Data Quality Elements

OLCI data derived from the CyAN project can be assessed through the lens of SWAMP Data quality indicators (DQIs), analogous to what is routinely evaluated for a variety of sample matrices and parameters collected during *in-situ* water quality and toxicity monitoring efforts. DQIs are quantitative measures and qualitative descriptors used to set limits of acceptable levels of data error. Here DQIs are defined using a set of elements related to data quality assessment. The measures include accuracy/bias, precision, representativeness, comparability, completeness, as well as sensitivity in the form of the detection limit. DQIs are used as a means to specify measurement quality objectives (MQOs) which, if achieved, will provide an indication that the resulting data are valid and expected to meet the project data quality objectives (EPA, 2002). This provides a process to set an acceptable amount of uncertainty for the data collected from a project, and ultimately, to assess project performance and confidence in the resulting data. This report will support development of MQOs that will be articulated in the future quality assurance project plan governing the use of satellite remote sensing data.

These data quality elements are:

- **Detection Limit** – A method to inform sensitivity of a measurement, is the lowest concentration of an analyte that a specified analytical procedure can reliably detect. Often called the minimum detection limit of an analytical procedure.

- **Accuracy/Bias** – The difference between an observed value and the "true" value (or known concentration) of the parameter being measured; bias is the first component of accuracy, which is the ability to obtain precisely a nonbiased (true) value. Sources of bias that can be introduced into analytical systems include systematic error, matrix interference, calibration, and contamination.

- **Precision** – The level of agreement among multiple measurements of the same characteristic; precision is the second component of accuracy.
- **Representativeness** – The degree to which the data collected accurately represents the population of interest (e.g., contaminant concentrations.)
- **Comparability** – The similarity of data from different sources included within individual or multiple data sets; the similarity of analytical methods and data from related projects across areas of concern.
- **Completeness** – The quantity of data that is successfully collected with respect to the amount intended in the experimental design.

OLCI imagery data, specifically CIcyano values and cyanobacterial Chl-a estimations as well as ESA sensor performance reports, were reviewed for these six elements in the sections below.

Inherent differences exist between point-based *in-situ* observations and raster-based, two-dimensional satellite data that limit comparisons to *in-situ* data. These six data quality elements are most related to data derived from field-based sampling; however, they provide a useful framework by which to evaluate satellite imagery and consistent with the SWAMP quality assurance manual requiring that applicable DQIs are evaluated for every measurement to inform confidence in the results. Three key caveats exist that are relevant to applying these data quality elements. First, there is an inherent spatial mismatch that exists between a point-based observation like a grab water sample and a spatially integrated observation of a pixel that covers a 300 x 300 m area. Given these spatial differences, perfect agreement is not expected. Next, imagery data integrates the visible portion of the water column, with the total integrated depth varying based on water clarity. While field measurements can also use an integrated water sampler to integrate depth, it is most common that the field collections involve surface water grabs. Lastly, there is often temporal variability between point-based observations and satellite imagery (on the order of minutes to hours). This can be a major driver of variation between these types of observations due to the highly variable nature of cyanobacteria, which can shift their positions within the water column throughout the day by modulating buoyancy and can be shifted across the water surface throughout the day by wind.

6.2 Detection Limit

Satellite data downloaded from the CyAN project is presented as digital numbers (DN) for each pixel, and DN = 1 - 253 is the range of values that indicate cyanobacteria detection. The lower detection limit of the OLCI sensor is difficult to quantify, however the best estimate provided by the CyAN project is that the lower detection limit for CIcyano is between 10,000 – 20,000

cells/mL (Coffer et al., 2021a). DNs of 1 – 253 are roughly equivalent to ~10-20,000 – 7,000,000 cells/mL. It should be noted though that the conversion of CIcyano to cell abundance data is currently a very rough estimation (more details in **Chapter 3**). The relatively high cell abundance equivalent of the lower end of the detection limit may make it hard to detect the beginning of a cyanobacterial bloom using satellite data. Therefore, low detections of cyanobacteria abundance may indicate that impacts to beneficial uses are already occurring in these waterbodies.

The detection of total Chl-a using the RE10 algorithm described in Chapter 3 is most effective beginning at concentrations > 1 µg/L (Wynne et al., 2022). Additionally, the conversion between CIcyano values to cyanobacterial Chl-a was shown to perform best beginning at *in-situ* Chl-a concentrations around 7 µg/L (Seegers et al., 2021).

6.3 Bias

The multidimensional nature of satellite imagery data makes it challenging to derive a "true value" by which to compare (see description of key differences above). For analytical chemistry systems, the true value is based on a standardized reference material for the analyte to be measured or a surrogate analyte; this approach is not feasible for satellite remote sensing. To measure bias for this system, one approach is to compare remotely sensed versus point-based field observations to ascertain if there is agreement between a field observation and a satellite image collected from a similar time and pixel location. This approach can serve to determine if satellite imagery is generally in agreement with a field observation. It is important to note, however, that field-based data from point sampling has its own bias that complicates these types of comparisons. For example, Trees et al. (1985) found that fluorometric measurements of Chl-a underestimated concentrations by an average of 39% compared to measurement via high-performance liquid chromatography in samples collected from three different oceanic regions. Despite this known bias associated with fluorometric Chl-a determinations, it is one of the most common approaches for analysis used today.

Field validation using *in-situ* radiometry can help to assess and cross validate if satellite data are within the published range and identify potential thresholds of increased false-positive risks but also are not perfect at generating a "true value" by which to compare. Mismatches between field and satellite-based sensors may also occur due to environmental and instrument variability, such as moving parcels of water or difference in instrument calibration, and do not necessarily indicate imprecise satellite measurements. Variability in the matchups between satellite and field radiometry data is an inherent issue with cyanobacteria blooms and thus field validation efforts (Tomlinson et al 2016).

Sentinel-3 satellite imagery was able to detect cyanobacteria blooms that occurred in Clear Lake (Lake County, CA) in a field validation study conducted by the CA State Water Boards in 2019 – 2020 (Lie et al., *in prep*). In this study, satellite data from CyAN performed better in bloom detection than field measurements as a majority (87%) of the field measurements indicated no bloom presence (i.e., DN = 0) despite macro- and microscopic confirmation of cyanobacteria in the field at the same time as the satellite overpass. Field radiometry generated DN were mostly 0 due to negative values of the exclusion criterion SS(665) in the calculation of CIcyano. Satellite data also failed to detect blooms in waterbodies other than Clear Lake (e.g., Lake San Antonio) despite microscopic observations of *Dolichospermum* spp. in water samples. Low cyanobacteria biomass could be a potential reason why satellite and field radiometry failed to detect blooms. Therefore, while the addition of the SS(665) criterion for CIcyano values was meant to reduce false-positives (i.e., CI value > 0 despite the absence of cyanobacteria), this study demonstrated that instances of false negative can occur, especially at lower cyanobacteria biomass. Further studies could be conducted to investigate a possible threshold in which SS(665) (and thus CIcyano and DN) becomes unreliable in detecting the presence of cyanobacteria. Another potential reason for the failure to detect blooms is the composition of the cyanobacteria community. The use of the SS(665) exclusion criterion is based on the observation of a large peak at 664 nm in *Microcystis* dominated water in an African dam (Matthews et al. 2012), so the inability for SS(665) to indicate the presence of cyanobacteria in the study may be due to differences in optical properties between *Dolichospermum* and other cyanobacteria genera compared to those of *Microcystis*. This highlights the need for additional work to better resolve the performance of the CIcyano observations in systems with variable cyanobacterial community composition. Nevertheless, in cases of the remaining 13% of field generated DN with values > 0, comparison with CyAN DN showed generally good agreement (Field validation report; Fig. 10).

6.4 Precision

Two Sentinel-3 satellite platforms are currently in orbit, each with an identically designed OLCI sensor, thus it is possible to assess precision by assessing differences across repeated measures of the same waterbody location by each individual sensor onboard Sentinel-3A and Sentinel-3B. The ESA conducted a tandem flight configuration to cross compare the in-flight performance of each sensor, OLCI-A and OLCI-B, from early June to mid-October following the launch of Sentinel-3B in 2018. A tandem flight configuration means each satellite can take images of the same location roughly 30 seconds apart from the same viewing angle, which allowed for exploration of potential differences between the platforms once in flight (Lamquin et al., 2020a). From these observations, it was determined that a radiometric bias of 1-2% existed between the two sensors (Lamquin et al., 2020a). Given this minor difference between sensors,

Hammond et al. (2020) examined how individual sensor uncertainty may influence trend and spatial pattern analysis during a period of tandem flight configuration. This study found that differences in sensor uncertainty resulted in minimal differences in trends of Chl-a in coastal and open ocean environments, and that suggested future trend analysis using both satellite sensor platforms should yield consistent results with continued sensor performance (Hammond et al., 2020).

Following the tandem flight phase, Sentinel-3 flight pattern was shifted from the tandem configuration to their current orbit phasing, where the Sentinel-3B satellite has a 140° separation from the Sentinel-3A satellite. This shift allows for the current near daily temporal resolution of the Sentinel-3 mission. A progressive drift phase was utilized to slowly make this shift and allow for additional comparisons between the OLCI-A and OLCI-B sensors across a range of viewing angles (Lamquin et al., 2020a). During this drift phase, the small differences in sensors and the downstream impacts on data products like Chl-a were examined and were found to be minimal (Lamquin et al., 2020a). The temporal stability of Chl-a cross-calibrations between the OLCI-A and OLCI-B was also assessed and was found to be stable between a roughly 1.5-year period following the drift phase (Lamquin et al., 2020b). Similar comparisons of CIcyano have not been reported. The differences between sensors continues to be monitored by the ESA and monthly assessments of OLCI performance and stability are conducted to ensure continued data quality. The most recent report from the ESA found negligible variation between sensor performances (ESA, 2024). Two Level 2 products are evaluated by the ESA routinely (the Green Instantaneous Fraction of Absorbed Photosynthetically Available Radiation and the OLCI Terrestrial Chlorophyll Index) that allow comparison between Sentinel-3A and Sentinel-3B. When comparing the outputs of each individual satellite for each of these metrices, correlations between each satellite ranged from 0.90 – 0.95, bias ranged from 0.01-0.03 and normalized root mean square error ranged from 0.07-0.18 in 2023 (ESA, 2024).

6.5 Representativeness

Each pixel of satellite data represents the average surface condition within the pixel. This is despite potential heterogeneity within the pixel, as is a common source of error for field-based measurements of cyanobacterial blooms which can have high spatial variability within a small area. As described above, each pixel of satellite data also integrates measurements over the depth of the photic zone, which varies based on water clarity conditions. Cyanobacterial biomass can also vary across depth based on regulation of buoyancy (ability to change position throughout the day); Graham et al. (2008) described potential water column distributions of cyanobacteria, some of which are more readily detectable by remote sensing methods than others. Field based survey methods also can have the same challenges to measure water

column distributions of cyanobateria due to constraints of resources, equipment, and access to open water.

Another consideration of representativeness is how cyanobacterial cells are spatially distributed. Satellite data are well suited to capture the full bloom intensity if cells are concentrated at the surface (e.g., *Microcystis* blooms under calm water conditions), but sometimes blooms may be spread out vertically within the water column (e.g., *Planktothrix* blooms or *Microcystis* bloom homogenization via wind action). Cyanobacteria blooms also often accumulate along the shore due to wind and wave action, but satellite measurements are limited to open water areas only due to issues with adjacency effects and mixed pixels closer to shore. Altogether, the constraints may result in underestimation of the true cyanobacteria abundance due to the omission of shorelines and particularly during times of well mixed water columns.

Despite this potential for *underestimation* due to the limitation of imagery data, studies of cyanobacterial blooms using satellite imagery have aligned well with field-based observations of bloom phenology. Patterns in weekly near-surface, open water cyanobacterial bloom occurrence were studied using both MERIS and OLCI imagery observations of more than 2000 large lakes and reservoirs nationwide. This study, which used imagery data collected between 2008-2011 and 2017-2018, demonstrated that weekly cyanobacterial bloom occurrence patterns on the national scale followed the anticipated temporal pattern of freshwater blooms. Overall, the number of waterbodies experiencing blooms each week increased throughout the calendar year, reaching a peak in the late summer and early fall, and decreasing with the onset of winter (Coffer et al., 2020). This demonstrates that over a large spatial scale, satellite imagery data can typically reproduce previously reported bloom phenology (Jöhnk et al., 2007).

6.6 Comparability

Several recent studies have been published that relate CIcyano values to a series of *in-situ* indicators of cyanobacterial blooms that help to understand how imagery data related to these observations. CIcyano values have been cross validated against the presence of a cyanotoxin class named microcystins (as an indicator of blooms presence) in 281 lakes across eleven states across the United States (Mishra et al., 2021). These comparisons were made using both the MERIS and OLCI sensors for the period of 2005-2011 and 2016-2019. Same day comparisons of CIcyano values and microcystins observations demonstrated that imagery data had an overall accuracy of 84% for detecting blooms when using toxin levels as an indicator of bloom presence (Mishra et al., 2021). In another study, eleven lakes with drinking water intakes across six different states (including California) were assessed for cyanobacterial bloom frequency using OLCI imagery between 2018 and 2019 (Coffer et al., 2021a). Satellite observations were

compared to 92 qualitative, visual observations of algal bloom presence near source water intakes (e.g., within 900 m). This comparison showed a 94% agreement between the imagery data and the visual observation data (Coffer et al., 2021). Schaeffer et al. (2018) and Whitman et al. (2022) both conducted comparisons between satellite imagery data and state issued cyanobacteria advisories and bloom reports. Schaeffer et al. (2018) compared OLCI imagery data to 25 different state advisories issued in six different states (including California) in 2017. State advisories, although most utilizing somewhat variable bloom definitions, were used as a general indication of *in-situ* bloom presence. Satellite data correctly identified bloom events that co-occurred with these advisories (Schaeffer et al., 2018). In a subsequent study, Whitman et al. (2022) examined the presence-absence agreement between reported cyanobacterial bloom events (a Natural Resources Defense Council database) and state advisories (state event response databases) to MERIS and OLCI imagery data. The bloom event database used for cross validation contained 1,343 events from 210 lakes across 23 states (including California), while the state advisory database included 160 advisories from 87 lakes across 11 states. Both cross validation datasets considered the period of 2008-2019. This study found a 69% agreement rate with state issued advisories and a 60% agreement rate with the bloom event data (Whitman et al., 2022). Collectively, these studies report agreement rates between 60% and 94% between a variety of direct and indirect field-based indicators (representing multiple lines of evidence) of cyanobacterial blooms and Clcyano values.

6.7 Completeness

The completeness of satellite data for resolvable waterbodies is affected by weather conditions such as cloud coverage or sun glint (Figure 6.1). Reflectance from ice and snow can also influence satellite data and pixels with snow and ice coverage should be removed from analysis (See section 6.4.2). Therefore, the amount of satellite data available in areas or seasons with high cloud and snow coverage tends to be reduced. Coffer et al. (2020) presented a summary of seasonal variations in bloom occurrence and lake resolvability across the United States. In the southwest region, which included California and Nevada, approximately 10%-15% of the 101 lakes included in the study were resolvable throughout the entire year depending on the year. However, during the late spring through late fall period, nearly 100% of lakes were resolvable, even after ice pixel removal (Coffer et al., 2020). Adopting a process to identify when satellite remote sensing data is not complete (or available) due to these environmental conditions is a method to avoid the error associated with incomplete data.

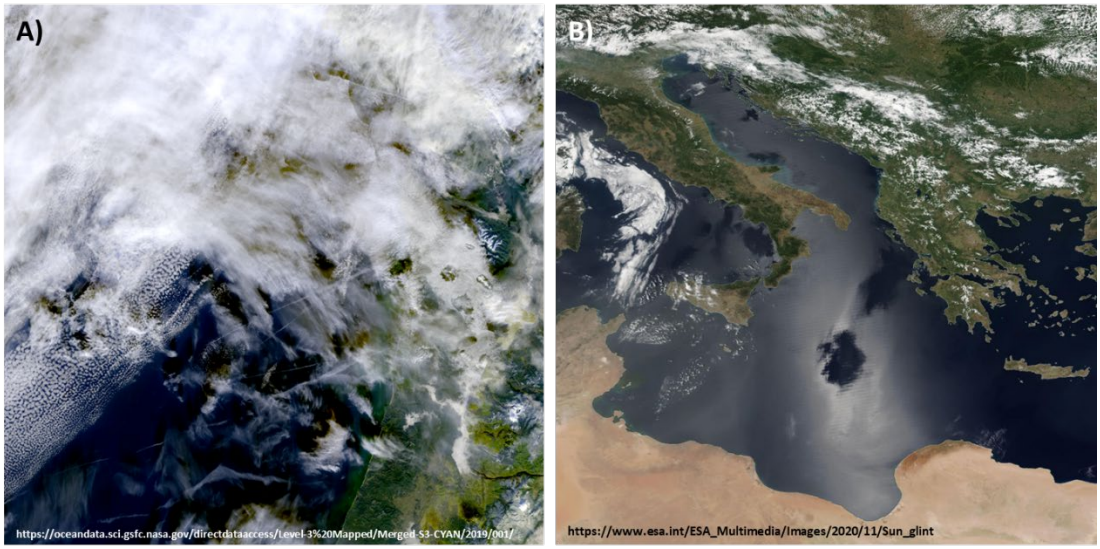

Figure 6.1: Examples of A) cloud coverage (http://oceandata.sci.gsfc.nasa.gov/getfile/L2019001.L3m_DAY_CYANTC_tc_CYAN_CONUS_300m_1_1.tif) and B) sun glint (https://www.esa.int/ESA_Multimedia/Images/2020/11/Sun_glint) from true color satellite images. Picture of sun glint is taken from the European Space Agency website.

7. Future Recommendations to Enhance and Streamline Applications

The FHAB Program intends to expand the use of satellite imagery for FHAB monitoring and assessment. Their long-term goal is to expand the application of FHAB remotely sensed data to support Water Board management decisions beyond a screening tool for event response (Table 7.1). This report provides the initial documentation of standard procedures and data quality to support the provisioning of high-quality data and document uncertainties.

Table 7.1: FHAB Program vision for Water Board-specific applications of FHAB satellite remote sensing. Examples of each application is provided in the description column.

FHAB Program's Proposed Applications to relevant Water Board Programs	Description
Waterbody screening for event response	Evaluating waterbodies for immediate FHAB response to implement the Voluntary Response Plan to inform public protection
Assessment	Evaluating status and trends at the waterbody, watershed, regional, and statewide level
Evaluating Impairments	Supporting decisions on impairments due to FHABs
Informing Permit Requirements	Where waterbodies are identified to be at risk of FHABs, supporting the development of permit requirements to prevent FHABs and/or incorporation of FHAB-specific monitoring requirements
Evaluation of TMDL Implementation, Restoration, or Enforcement Actions	Evaluating effectiveness of actions taken to mitigate FHABs in a waterbody

7.1 Conclusions and Future Recommendations

This report outlines a streamlined robust approach for Water Board and partner water quality applications by: 1) Identifying the remote sensing indicators and metrics that correspond to specific management questions; 2) establishing best practices for FHAB data processing, building on the considerable foundation of CyAN published literature and expert assistance; 3) identifying how these indicators and metrics would be calculated and visualized to assess FHAB

status and trends in single lakes versus groups of lakes, available to the community through well documented, open source coding; and 4) review of the literature describing what is understood about satellite remote sensing data quality. Implementing the approaches described in this report provides a tractable approach to assess the status and trends of large lakes and other waterbodies, which is particularly valuable since very little FHAB related data are available in these systems due to the lack of ambient programs in the majority of these lakes.

Below, we identify multiple recommendations to further streamline the adoption of satellite remote sensing for FHAB assessment. These recommendations fall with five major types of activities, discussed briefly in the sections below.

A. Integrate additional satellite platforms that can greatly expand the extent of waterbodies characterized

B. Develop consensus on action levels for remotely sensed FHAB data

C. Continue to develop quality assurance studies and documentation for application of satellite remote sensing to Water Board water quality programs.

D. Enhance data communication, accessibility, visualizations, and reporting that can increase the utility of remote sensed data for the Water Boards and their partners.

E. Integrate satellite remote sensed data into FHAB driver assessments.

A. Collaborate with CyAN Project et al. partners to integrate additional satellite platforms (e.g. Sentinel 2 MSI imagery) to greatly expand the extent of waterbodies characterized

Higher spatial resolution platforms do exist and are envisioned as a valuable data stream in the future for FHAB monitoring and assessment. Of greatest interest is the Multispectral Instrument (MSI) onboard the Sentinel-2 satellite constellation. This sensor offers a 30-meter spatial resolution which will allow for a majority of waterbodies in California to be monitored remotely. MSI generates imagery with a different spectral signature; therefore, these products cannot be used to calculate a CIcyano value. Methods do exist to use MSI imagery to estimate Chl-a concentrations, which could provide data about overall algal biomass, but a cyanobacteria specific index equivalent to the CIcyano index is not currently available. Nonetheless, MSI and other high resolution satellite imagery products offer promising additions to FHAB monitoring. One of the major limitations for including MSI data products in this current effort is that Level-3 data products from Sentinel-2 are not currently routinely available in a similar ready to use format as the Sentinel-3 products. Due to the increased spatial resolution, a significantly larger amount of data is generated by this platform compared to OLCI imagery. Significant

investments would be needed to manage and use this data if this was done directly by the state. The CyAN project has kicked off efforts in 2024 to build similar infrastructure as currently available for OLCI data products. Thus, future use of Sentinel-2 and other higher resolution data products were kept in mind throughout the development of this report, and many of the recommendations for data processing and analysis in this report are envisioned to be scalable to these future high resolution data products.

B. Develop consensus on FHAB action levels for satellite remote sensing to support water quality management decisions

To improve utility of remotely sensed data for management decisions, thresholds should be defined that link satellite data (CIcyano and chlorophyll-*a*) to risk of exceeding thresholds of *in-situ* pigment data (chlorophyll-a and phycocyanin) that impair beneficial uses. The infrastructure to conduct this data match up process is readily available since in recent years, approaches have been developed to streamline matching co-located *in-situ* measurements with satellite observations (Ross et al. 2019), making this an accessible approach. As additional in-situ data are developed this relationship can be refined over time. As an example of such studies, one could investigate the application of the thresholds currently recommended in this report (e.g., bloom alert threshold based on WHO 2021 guidance) and determine how well this threshold aligns with in-situ observations.

C. Continue to develop quality assurance studies and documentation for applying satellite remote sensing to Water Board water quality programs

Within this recommendation, two specific actions are recommended. First, the SWAMP program should develop a generalized quality assurance project plan (QAPP) for application of satellite remote sensing to Water Board water quality programs. This document provides the scientific review with which to accelerate the development of that QAPP. The QAPP should identify which decisions would be supported and the action levels (Recommendation B above).

Second, SWAMP should continue to validate and assess remote sensing products with available *in-situ* data to further evaluate algorithm performance. Field sampling could be aligned for dates with satellite overpasses to increase the *in-situ* data available for validation. These data are best collected in collaboration with other academic and research partners. Therefore, a specific strategy, SOP, and data management system for satellite field verification should be developed, so that external partners can help participate in collecting data and submit it to a common database. These efforts are envisioned to continue to ground-truth satellite data, improve the CA-specific algorithm and overall data quality characterization. These efforts should be conducted in close collaboration with this expert workgroup and the strategy and subsequent SOP should be published and widely distributed among partners in the state.

D. Increase the utility of remote sensed data for the Water Boards and their partners.

To expand the user community for remotely sensed FHAB data, it is essential to build data systems and decisions support incrementally even at moderate FHAB monitoring program funding levels to encourage strong partnerships and rapid dissemination of FHAB monitoring data. To achieve this, we recommend developing the vision, including key data sources, data visualizations and graphical user interface (GUI) functionality, for each type of decision support, through interactions with intended user groups (Waterboard or external targeted FHAB monitoring program partners). Two near-term actions are highlighted:

1) *Integrating satellite remote sensing into regular reporting on FHAB status and trends.*

Annual or biannual reports could be published by SWAMP on a 1- to 2-year interval and would summarize FHAB status, trends, and drivers from imagery data of visible waterbodies on a regional, watershed, and statewide basis.

2) *Integrate data metrics and summaries into a FHAB Satellite visualization tool to allow for functionalities such as comparison of a specific lake to the larger population of lakes and for download of data metrics into an easily digestible format such as an excel or csv file.*

The State Water Boards has already made strategic investments to capitalize on federally curated FHAB remote sensing products for large lakes and provides these data through a Californian FHAB satellite portal (fhab.sfei.org). The metrics and workflows, including the R Shiny app for status assessment, will be incorporated into the routine workflows broadly available to the portal. A strategic vision for the portal, and incremental development steps should be further identified, especially as new remote sensing options become available.

E. Integrate satellite remote sensed data in FHAB drivers assessment

To expand the utility of satellite remotely sensed data resulting from these best practices, we recommend this data as routine data source to inform FHAB driver assessments. The FHAB Monitoring Strategy (Smith et al. 2021) provided a comprehensive review of external and internal drivers and potential data sources. We recommended assessing the status of data that can be used routinely used to assess drivers, identifying metrics, workflows required to investigate the relationship between status and drivers, including the assessment of data quality. Leveraging remote sensing in a standardized approach can support elements of the FHAB Monitoring Strategy to holistically assess the drivers, the occurrence and inform management decisions to address HABs.

REFERENCES

Backer, L. C., Landsberg, J. H., Miller, M., Keel, K., and Taylor, T. K. (2013). Canine cyanotoxin poisonings in the United States (1920s–2012): Review of suspected and confirmed cases from three data sources. *Toxins* 5, 1597–1628.

Burr, G. S., Wolters, W. R., Schrader, K. K., and Summerfelt, S. T. (2012). Impact of depuration of earthy-musty off-flavors on fillet quality of Atlantic salmon, Salmo salar, cultured in a recirculating aquaculture system. *Aquacultural Engineering* 50, 28–36.

Chorus, I., and Bartram, J. eds. (1999). *Toxic cyanobacteria in water: a guide to their public health consequences, monitoring, and management*. London ; New York: E & FN Spon.

Chorus, I., and Welker, M. (2021). *Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management*. Taylor & Francis.

California State Water Resources Control Board. (2021, 2022, 2023). Water Code Section 13181(a) Freshwater and Estuarine Harmful Agal Bloom Annual Reports. Available at: https://www.waterboards.ca.gov/water_issues/programs/swamp/freshwater_cyanobacteria.html

Clark, J. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., Johnston, J. M., Ignatius, A. R., et al. (2017). Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. *Ecological Indicators* 80, 84–95. doi: 10.1016/j.ecolind.2017.04.046

Clerc, S., Donlon, C., Borde, F., Lamquin, N., Hunt, S. E., Smith, D., et al. (2020). Benefits and Lessons Learned from the Sentinel-3 Tandem Phase. *Remote Sensing* 12, 2668. doi: 10.3390/rs12172668

Coffer, M. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., and Salls, W. B. (2020). Quantifying national and regional cyanobacterial occurrence in US lakes using satellite remote sensing. *Ecological Indicators* 111, 105976. doi: 10.1016/j.ecolind.2019.105976

Coffer, M. M., Schaeffer, B. A., Foreman, K., Porteous, A., Loftin, K. A., Stumpf, R. P., et al. (2021a). Assessing cyanobacterial frequency and abundance at surface waters near drinking water intakes across the United States. *Water Research* 201, 117377. doi: 10.1016/j.watres.2021.117377

Coffer, M. M., Schaeffer, B. A., Salls, W. B., Urquhart, E., Loftin, K. A., Stumpf, R. P., et al. (2021b). Satellite remote sensing to assess cyanobacterial bloom frequency across the United

States at multiple spatial scales. *Ecological Indicators* 128, 107822. doi: 10.1016/j.ecolind.2021.107822

ESA (2024). Sentinel-3 optical Annual Performance Report – Year 2023. Sentinel Online: European Space Agency. Available at: https://sentiwiki.copernicus.eu/__attachments/1681931/OMPC.ACR.APR.004%20-%20S3%20Optical%20Annual%20Performance%20Report%202023%20-%201.1.pdf?inst-v=7e368646-a179-477f-af62-26dcc645dd8a (Accessed October 21, 2024).

Glibert, P. M., Landsberg, J. H., Evans, J. J., Al-Sarawi, M. A., Faraj, M., Al-Jarallah, M. A., et al. (2002). A fish kill of massive proportion in Kuwait Bay, Arabian Gulf, 2001: the roles of bacterial disease, harmful algae, and eutrophication. *Harmful Algae* 1, 215–231. doi: 10.1016/S1568-9883(02)00013-6

Gons, H. J., Hakvoort, H., Peters, S. W., and Simis, S. G. (2005). "Optical detection of cyanobacterial blooms," in *Harmful cyanobacteria*, (Springer), 177–199.

Graham, N. J. D., Wardlaw, V. E., Perry, R., and Jiang, J.-Q. (1998). The significance of algae as trihalomethane precursors. *Water Science and Technology* 37, 83–89. doi: 10.1016/S0273-1223(98)00013-4

Graham, J. L., Loftin, K. A., Ziegler, A. C., and Meyer, M. T. (2008). Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs. US Geological Survey.

Hammond, M. L., Henson, S. A., Lamquin, N., Clerc, S., and Donlon, C. (2020). Assessing the Effect of Tandem Phase Sentinel-3 OLCI Sensor Uncertainty on the Estimation of Potential Ocean Chlorophyll-a Trends. *Remote Sensing* 12, 2522. doi: 10.3390/rs12162522

Hirsch, R. M., Slack, J. R., and Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* 18, 107–121. doi: 10.1029/WR018i001p00107

Hirsch, R. M. (1988). Statistical Methods and Sampling Design for Estimating Step Trends in Surface-Water Quality. *J American Water Resour Assoc* 24, 493–503. doi: 10.1111/j.1752-1688.1988.tb00899.x

Howard, M. D. A., Kudela, R. M., Hayashi, K., Tatters, A. O., Caron, D. A., Theroux, S., et al. (2021). Multiple co-occurring and persistently detected cyanotoxins and associated cyanobacteria in adjacent California lakes. *Toxicon* 192, 1–14. doi: 10.1016/j.toxicon.2020.12.019

Howgate, P. (2004). Tainting of farmed fish by geosmin and 2-methyl-iso-borneol: a review of sensory aspects and of uptake/depuration. *Aquaculture* 234, 155–181.

Kudela, R. M. (2011). Characterization and deployment of Solid Phase Adsorption Toxin Tracking (SPATT) resin for monitoring of microcystins in fresh and saltwater. *Harmful Algae* 11, 117–125.

Jöhnk, K. D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P. M., and Stroom, J. M. (2008). Summer heatwaves promote blooms of harmful cyanobacteria. *Global Change Biology* 14, 495–512. doi: 10.1111/j.1365-2486.2007.01510.x

Lamquin, N., Déru, A., Clerc, S., Bourg, L., and Donlon, C. (2020a). OLCI A/B Tandem Phase Analysis, Part 2: Benefits of Sensors Harmonisation for Level 2 Products. *Remote Sensing* 12, 2702. doi: 10.3390/rs12172702

Lamquin, N., Bourg, L., Clerc, S., and Donlon, C. (2020b). OLCI A/B Tandem Phase Analysis, Part 3: Post-Tandem Monitoring of Cross-Calibration from Statistics of Deep Convective Clouds Observations. *Remote Sensing* 12, 3105. doi: 10.3390/rs12183105

Legleiter, C. J., King, T. V., Carpenter, K. D., Hall, N. C., Mumford, A. C., Slonecker, T., et al. (2022). Spectral mixture analysis for surveillance of harmful algal blooms (SMASH): A field-, laboratory-, and satellite-based approach to identifying cyanobacteria genera from remotely sensed data. *Remote Sensing of Environment* 279, 113089. doi: 10.1016/j.rse.2022.113089

Lunetta, R. S., Schaeffer, B. A., Stumpf, R. P., Keith, D., Jacobs, S. A., and Murphy, M. S. (2015). Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. *Remote Sensing of Environment* 157, 24–34. doi: 10.1016/j.rse.2014.06.008

Mehinto, A.C., Smith, J., Wenger, E., Stanton, B., Linville, R., Brooks, B.W., Sutula, M.A., Howard, M.D., 2021. Synthesis of ecotoxicological studies on cyanotoxins in freshwater habitats–Evaluating the basis for developing thresholds protective of aquatic life in the United States. Science of The Total Environment 795, 148864.

Miller, M. A., Kudela, R. M., Mekebri, A., Crane, D., Oates, S. C., Tinker, M. T., et al. (2010). Evidence for a Novel Marine Harmful Algal Bloom: Cyanotoxin (Microcystin) Transfer from Land to Sea Otters. *PLoS ONE* 5, e12576. doi: 10.1371/journal.pone.0012576

Mishra, S., Stumpf, R. P., Schaeffer, B. A., Werdell, P. J., Loftin, K. A., and Meredith, A. (2019). Measurement of Cyanobacterial Bloom Magnitude using Satellite Remote Sensing. *Sci Rep* 9, 18310. doi: 10.1038/s41598-019-54453-y

Mishra, S., Stumpf, R. P., Schaeffer, B., Werdell, P. J., Loftin, K. A., and Meredith, A. (2021). Evaluation of a satellite-based cyanobacteria bloom detection algorithm using field-measured microcystin data. *Science of The Total Environment* 774, 145462. doi: 10.1016/j.scitotenv.2021.145462

National Research Council (NRC) (2000). *Clean coastal waters: understanding and reducing the effects of nutrient pollution*. Washington, DC: National Academies Press.

Paerl, H.W., Otten, T.G., Kudela, R., 2018. Mitigating the Expansion of Harmful Algal Blooms Across the Freshwater-to-Marine Continuum. Environ. Sci. Technol. 52, 5519–5529. https://doi.org/10.1021/acs.est.7b05950

Papenfus, M., Schaeffer, B., Pollard, A. I., and Loftin, K. (2020). Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. *Environ Monit Assess* 192, 808. doi: 10.1007/s10661-020-08631-5

Puschner, B., Hoff, B., and Tor, E. R. (2008). Diagnosis of Anatoxin-a Poisoning in Dogs from North America. *J VET Diagn Invest* 20, 89–92. doi: 10.1177/104063870802000119

Plummer, J. D., and Edzwald, J. K. (2001). Effect of Ozone on Algae as Precursors for Trihalomethane and Haloacetic Acid Production. *Environ. Sci. Technol.* 35, 3661–3668. doi: 10.1021/es0106570

Reinl, K.L., Harris, T.D., North, R.L., Almela, P., Berger, S.A., Bizic, M., Burnet, S.H., Grossart, H., Ibelings, B.W., Jakobsson, E., Knoll, L.B., Lafrancois, B.M., McElarney, Y., Morales-Williams, A.M., Obertegger, U., Ogashawara, I., Paule-Mercado, M.C., Peierls, B.L., Rusak, J.A., Sarkar, S., Sharma, S., Trout-Haney, J.V., Urrutia-Cordero, P., Venkiteswaran, J.J., Wain, D.J., Warner, K., Weyhenmeyer, G.A., Yokota, K., 2023. Blooms also like it cold. Limnol Oceanogr Letters 8, 546–564. https://doi.org/10.1002/lol2.10316

Robin, J., Cravedi, J.-P., Hillenweck, A., Deshayes, C., and Vallod, D. (2006). Off flavor characterization and origin in French trout farming. *Aquaculture* 260, 128–138.

Schaeffer, B. A., Bailey, S. W., Conmy, R. N., Galvin, M., Ignatius, A. R., Johnston, J. M., et al. (2018). Mobile device application for monitoring cyanobacteria harmful algal blooms using Sentinel-3 satellite Ocean and Land Colour Instruments. *Environmental Modelling & Software* 109, 93–103. doi: 10.1016/j.envsoft.2018.08.015

Schaeffer, B. A., Urquhart, E., Coffer, M., Salls, W., Stumpf, R. P., Loftin, K. A., et al. (2022). Satellites quantify the spatial extent of cyanobacterial blooms across the United States at multiple scales. *Ecological Indicators* 140, 108990. doi: 10.1016/j.ecolind.2022.108990

Schaeffer, B. A., Reynolds, N., Ferriby, H., Salls, W., Smith, D., Johnston, J. M., et al. (2024). Forecasting freshwater cyanobacterial harmful algal blooms for Sentinel-3 satellite resolved U.S. lakes and reservoirs. *Journal of Environmental Management* 349, 119518. doi: 10.1016/j.jenvman.2023.119518

Seegers, B. N., Werdell, P. J., Vandermeulen, R. A., Salls, W., Stumpf, R. P., Schaeffer, B. A., et al. (2021). Satellites for long-term monitoring of inland U.S. lakes: The MERIS time series and application for chlorophyll-a. *Remote Sensing of Environment* 266, 112685. doi: 10.1016/j.rse.2021.112685

Shi, K., Zhang, Y., Qin, B., and Zhou, B. (2019). Remote sensing of cyanobacterial blooms in inland waters: present knowledge and future challenges. *Science Bulletin* 64, 1540–1556.

Smith, J., Eggleston, E., Howard, M. D. A., Ryan, S., Gichuki, J., Kennedy, K., et al. (2023). Historic and recent trends of cyanobacterial harmful algal blooms and environmental conditions in Clear Lake, California: A 70-year perspective. *Elem Sci Anth* 11, 00115. doi: 10.1525/elementa.2022.00115

Smith, J., Sutula, M., Bouma-Gregson, K., and Van Dyke, M. (2021). California Water Boards' Framework and Strategy for Freshwater Harmful Algal Bloom Monitoring: Full Report with Appendices. Sacramento, CA: California State Water Resources Control Board.

Stewart, I., Seawright, A. A., and Shaw, G. R. (2008). "Cyanobacterial poisoning in livestock, wild mammals and birds–an overview," in *Cyanobacterial harmful algal blooms: state of the science and research needs*, (Springer), 613–637.

Sutula, M., Shultz, D., Smith, J., Huie, S., 2025. Statistical models supporting decisions on eutrophication targets for California lakes and reservoirs. (Technical No. 1424). Southern California Coastal Water Research Project, Costa Mesa, CA.

Tomlinson, M. C., Stumpf, R. P., Wynne, T. T., Dupuy, D., Burks, R., Hendrickson, J., et al. (2016). Relating chlorophyll from cyanobacteria-dominated inland waters to a MERIS bloom index. *Remote Sensing Letters* 7, 141–149. doi: 10.1080/2150704X.2015.1117155

Trees, C. C., Kennicutt II, M. C., and Brooks, J. M. (1985). Errors associated with the standard fluorometric determination of chlorophylls and phaeopigments. *Marine Chemistry* 17, 1–12.

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

Urquhart, E. A., Schaeffer, B. A., Stumpf, R. P., Loftin, K. A., and 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. doi: 10.1016/j.hal.2017.06.001

USEPA (2000a). Guidance for assessing chemical contaminant data for use in fish advisories. Washington, DC: U.S. Environmental Protection Agency.

USEPA (2000b). Guidance for Data Quality Assessment, practical methods for data analysis. Washington, DC: U.S. Environmental Protection Agency. EPA QA/G-9. <https://www.epa.gov/sites/default/files/2015-06/documents/g9-final.pdf>

USEPA (2002). Guidance for Quality Assurance Project Plans. Washington, DC: U.S. Environmental Protection Agency. EPA QA/G-5, EPA/240/R-02/009.

US EPA, 2021. Ambient Water Quality Criteria to Address Nutrient Pollution in Lakes and Reservoirs. Office of Water, US Environmental Protection Agency, Washington, DC.

Whitman, P., Schaeffer, B., Salls, W., Coffer, M., Mishra, S., Seegers, B., et al. (2022). A validation of satellite derived cyanobacteria detections with state reported events and recreation advisories across U.S. lakes. *Harmful Algae* 115, 102191. doi: 10.1016/j.hal.2022.102191

Wynne, T. T., Meredith, A., Briggs, T., Litaker, W., and Stumpf, R. P. (2018). Harmful Algal Bloom Forecasting Branch Ocean Color Satellite Imagery Processing Guidelines. doi: 10.25923/twc0-f025

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., and Dyble, J. (2010). Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data. *Limnology and Oceanography* 55, 2025–2036.

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Warner, R. A., Tester, P. A., Dyble, J., et al. (2008). Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. *International Journal of Remote Sensing* 29, 3665–3672. doi: 10.1080/01431160802007640

Wynne, T. T., Tomlinson, M. C., Briggs, T. O., Mishra, S., Meredith, A., Vogel, R. L., et al. (2022). Evaluating the Efficacy of Five Chlorophyll-a Algorithms in Chesapeake Bay (USA) for Operational Monitoring and Assessment. *JMSE* 10, 1104. doi: 10.3390/jmse10081104

Yuan, L.L., Pollard, A.I., 2015. Deriving nutrient targets to prevent excessive cyanobacterial densities in U.S. lakes and reservoirs. Freshwater Biology 60, 1901–1916. https://doi.org/10.1111/fwb.12620

8. APPENDIX A. CHLOROPHYLL-*A* ESTIMATION THROUGH RE10 ALGORITHM

8.1 Introduction

Chlorophyll-a (Chl-a) is the dominant pigment light absorption pigments in phytoplankton and is often used as a proxy to total phytoplankton biomass. Accurate phytoplankton Chl-a measurements can be obtained by extracting Chl-a and then measurements via a fluorometer, but this method requires a physical water sample. Optical methods via remote sensing for measuring Chl-a rely on the optical properties of Chl-a, which absorbs light primarily in the violet (430 nm) and red (662 nm) wavelengths. Various algorithms have been developed for estimating Chl-a from satellite imagery data, including various Ocean Color (OCx), ESA algal_1 and algal_2, and the Red Edge 2010 (RE10).

RE10 determines Chl-a concentrations by calculating the near-infrared:red ratio. The use of the red and near infrared (NIR) bands is more effective for inland and coastal waters than the blue and green bands used for open ocean algorithms, as the red and NIR bands are less sensitive to interference from colored dissolved organic matter (CDOM) and particles scattering. The RE10 algorithm was first described by Gilerson et al. (2010) and has been subsequently modified by Wynne et al. (2018).

8.2 Methods

8.2.1 RE10 Algorithm

The RE10 algorithm is based on the ratio of near-infrared (709 nm):red (665 nm) ratio (R2), both atmospherically corrected with an assumption of a coarse-mode maritime aerosol with dark water at 885 nm:

$$R2 = \frac{\rho_s(709) - \rho_s(885)}{\rho_s(665) - \rho_s(885)}$$

Where ρ_s is the top-of-atmosphere reflectance for specific bands.

Gilerson et al. (2010) derived a 2-band algorithm and a 3-band algorithm for the estimation of Chl-a, but subsequent work indicated that the 2-band algorithm performed better.

$$Chla = \left\{\frac{a_w(\lambda_1)R2 - a_w(\lambda_2)}{0.022}\right\}^{\frac{1}{p}}$$

Where a_w are absorption due to water for wavelengths λ_1 and λ_2, and p is a coefficient generated from the relationship between phytoplankton Chl-a specific absorption coefficient (a^*_{ph}) between $a^*_{ph}(665)$ and $a^*_{ph}(675)$ (Gilerson et al. 2010; Bramich et al 2021). Taking $a_w(665)$ = 0.4245 m^{-1} and $a_w(709)$ = 0.7864 m^{-1}, and p = 0.89, the equation is rewritten as:

$$Chla = [35.75 \times R2 - 19.30]^{1.124}$$

Wynne et al. (2021) then regionally adjusted the coefficient of 19.30 to 14.30 to account for waters off south Florida as Chl-a values were not available from the original equation. Therefore, the final equation for RE10 used in Wynne et al. (2022) as well as this study is:

$$Chla = [35.75 \times R2 - 14.30]^{1.124}$$

A minimum Chl-a value of 0.5 µg/L is assigned when the calculated RE10 Chl-a was < 0.5 (Wynne et al 2021).

8.2.2 Satellite imagery processing

7D max composite NetCDF (.nc) files for 2019 - 2023 were downloaded from the Inland Waters Data (ILW) project on NASA EarthData's Ocean Color Website (https://oceancolor.gsfc.nasa.gov/about/projects/inlandwaters/), and then projected to match the pixel grid of the CyAN data before calculating the Chl-a using the RE10 algorithm. The data was then subsequently masked for mixed land-water and ice pixels using automated masking and NSIDC 4 km ice coverage as reference (See Section 4.3). The resulting rasters of Chl-a values as pixels were finally processed using R scripts to calculate the annual bloom frequency and weekly bloom occurrence, bloom spatial extent and bloom magnitude metrics. These R scripts are the same R scripts for processing CyAN data, but slightly modified to account for different Chl-a vs. CI bloom thresholds.

Calculated Chl-a values for pixels ranged from 0.5 – 6.8 x 10^8 µg/L, but extremely high values of calculated Chl-a were rare. Less than 2.5% of all resolvable pixels had Chl-a values > 200 µg/L for the 5 years analyzed (2019 – 2023), so a maximum cap of 200 µg/L was set such that Chl-a values >200 µg/L were set at 200 µg/L. A bloom alert threshold was defined as 12 µg/L was set for subsequent bloom metrics analysis. This bloom threshold was based on the recently updated WHO guidelines that utilize Chl-a observations in a risk based alert framework for waterbodies (Chorus and Welker, 2021), and mimicking the bloom alert threshold for cyanobacterial Chl-a estimates.

8.2.3 Mismatches with CyAN data

There are some discrepancies in pixel data availability between the ILW and CyAN datasets. Some pixels from the ILW dataset do not have any data for the bands used in the RE10

algorithm available, despite the same pixels from CyAN having a DN (Figure 8.1A). On the other hand, data flagged by CyAN for land (DN = 254) or no data (DN = 255) were not necessarily flagged by the NetCDF, and may be included in the final Chl-a rasters. An example is the red pixel seen in Figure 8.1B, which indicates a pixel from CyAN (blue color) was missing due to land flagging from CyAN. Finally, despite attempts to make sure that the projection for the NetCDF data from ILW match the rasters from CyAN, there is a slight misalignment (< 1 m) between the two (Figure 8.1C). Such misalignment might lead to differences in the pixels included in mixed and ice pixels masking. Together, these issues led to differences in the total number of pixels processed and analyzed between the two datasets (Figure 8.2), but a detailed investigation was not performed.

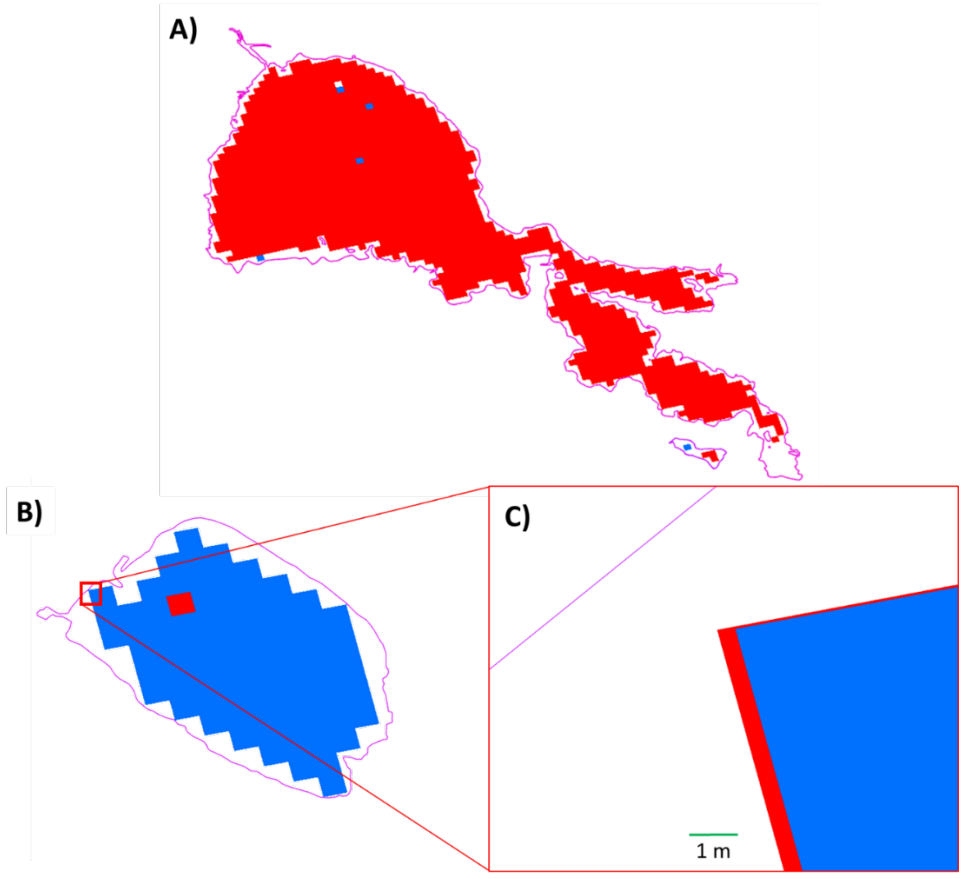

Figure 8.1: Comparison between the RE10 Chl-a pixels and the CyAN CI pixels using Clear Lake (A) and Lake Elsinore (B – C) during the first week of 2019 as an example. Blue color indicates CyAN pixels and red color indicates RE10 Chl-a pixels. A) Red RE10 Chl-a layer is on the top, so blue pixels indicate pixels with available data from CyAN that are not available for RE10 Chl-a. B) Blue CyAN layer is on the top, so red pixels indicate pixels with available data for RE10 Chl-a that is not available for CyAN. C) Close-up of B) that indicates the slight pixel misalignment between the RE10 Chl-a and CyAN pixels.

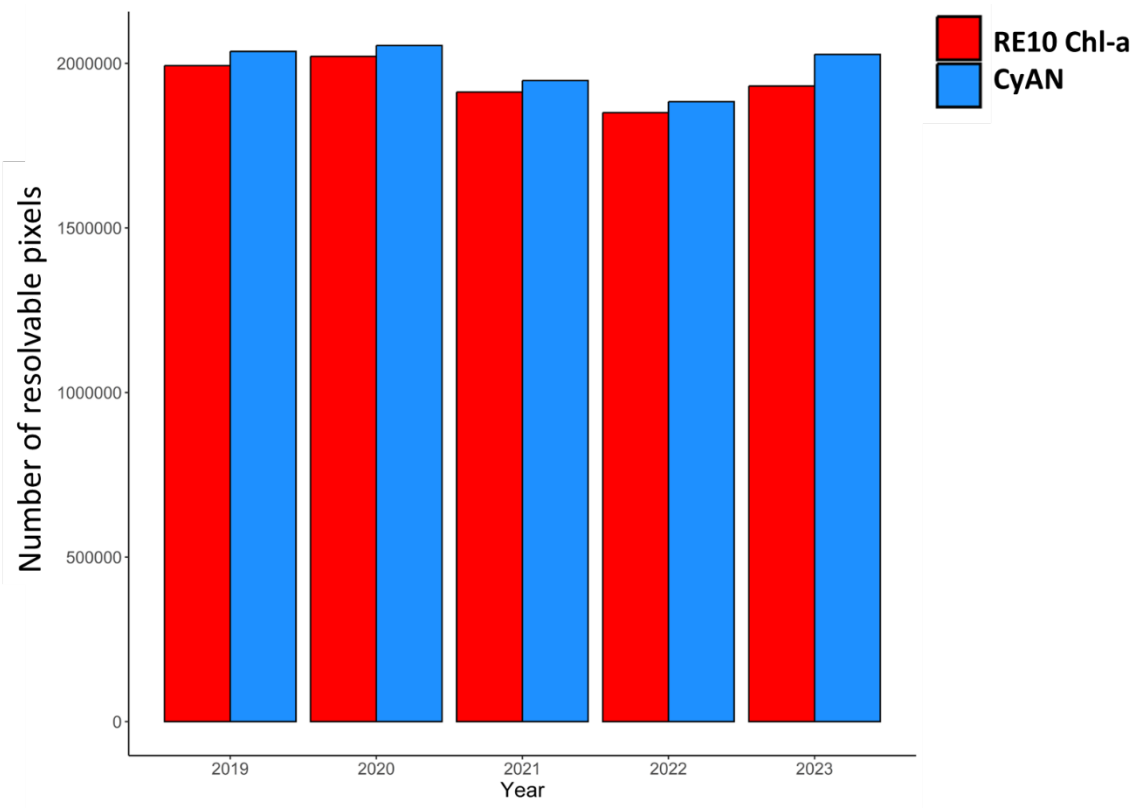

Figure 8.2: Comparison of the number of resolvable pixels between RE10 Chl-a (red bars) and CyAN (blue bars). Only waterbodies with 'Suitability for trend analysis'= Suitable were included (n = 191).

8.3 Comparison with CyAN CIcyano values

The RE10 Chl-a (total phytoplankton biomass) and CyAN CIcyano DN (a proxy for CIcyano) values for each pixel were compared to elucidate the relationship between CIcyano DN and RE10 Chl-a values. One goal in particular was to understand the relationship between the RE10 Chl-a estimates of the WHO bloom alert level of 12 µg/L (Chorus and Welker, 2021) compared to CIcyano bloom alert level (DN = 132). Bloom detection for RE10 Chl-a estimates was defined as 5 µg/L, and as DN > 0 for CIcyano.

Results indicated a generally logarithmic growth curve-type relationship (Figure 8.3), in which there is a linear portion of DN (< 210) and RE10 Chl-a (< 40 µg/L) and a plateau portion in which DN remains similar (~210 - 220) whereas RE10 Chl-a increased (> 40 µg/L). Figure 8.3. shows data points for pixels that have values for both CIcyano DN and RE10 Chl-a in 2023 as an example. Data points for all years (2019 – 2023) were not individually plotted due to the large amount of data points (~9.8 million), but results for each of the 5 years are similar to the results for 2023.

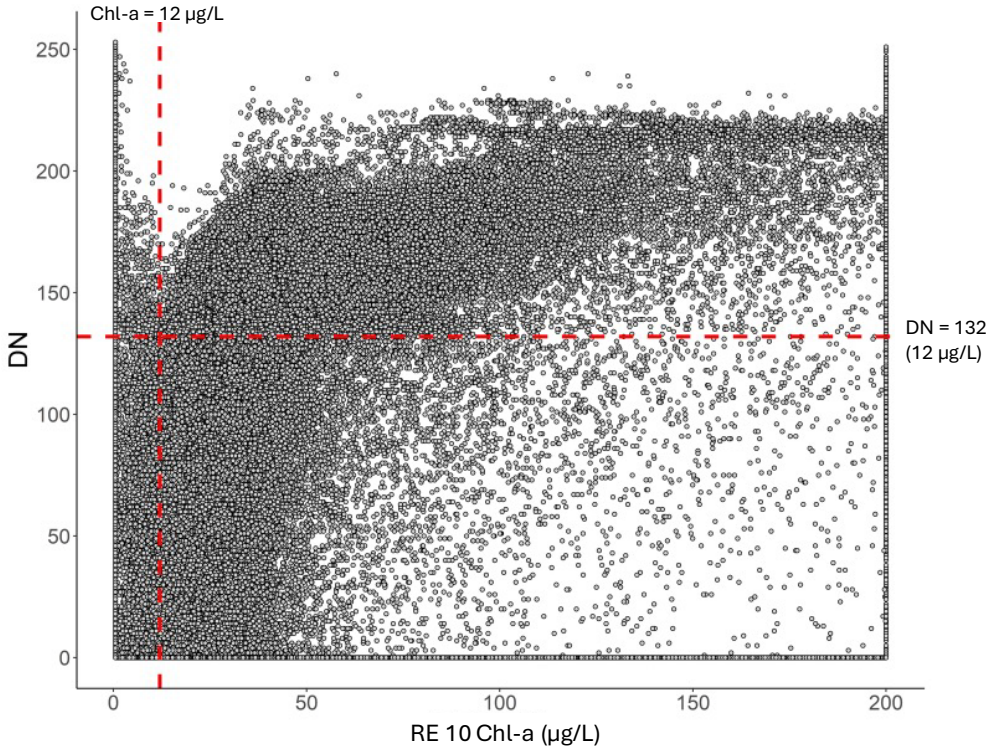

Figure 8.3: Comparison between individual pixel CIcyano DN and RE10 Chl-a values for paired pixels in 2023.

Only 6% of pixels had agreement for bloom detection classifications and 3% of pixels for bloom alert classifications between RE10 Chl-a and CIcyano DN biomass estimate comparisons (Table 8.1, Table 8.2). Agreement between no bloom classifications for each type of biomass estimation approach was slightly better with corresponding no bloom detection classifications and no bloom alert classifications occurring in 10% and 35% of pixels, respectively. More often than not, however RE10 Chl-a indicated a bloom detection classification (83% of pixels) or bloom alert classification (62% of pixels) when CIcyano DN values for the same pixels indicated these same bloom classifications were not present (Table 8.1, Table 8.2).

Table 8.1: Confusion matrix for bloom detection occurrences to evaluate agreement between CIcyano and chlorophyll *a* calculated from OLCI satellite imagery data from 2019 – 2023. Numbers indicate the number of pixels that fit into each scenario, and percentages in parentheses indicate the percentage of each scenario against the total number of pixels with paired CIcyano and chlorophyll *a* values (n = 9,700,830).

	Chl-a bloom detection	Chl-a no bloom detection
DN bloom detection	621,373 (6%)	70,742 (1%)
DN no bloom detection	8,050,228 (83%)	958,487 (10%)

Table 8.2: Confusion matrix for bloom alert occurrences to evaluate agreement between CIcyano and chlorophyll *a* calculated from OLCI satellite imagery data from 2019 – 2023. Numbers indicate the number of pixels that fit into each scenario, and percentages in parentheses indicate the percentage of each scenario against the total number of pixels with paired CIcyano and chlorophyll *a* values (n = 9,700,830).

	Chl-a bloom alert	Chl a no bloom alert
DN bloom alert	272,957 (3%)	56,955 (1%)
DN no bloom alert	6,018,645 (62%)	3,352,273 (35%)

8.4 Conclusion

We initiated procedures for estimating Chl-a values from OLCI data using the RE10 algorithm, but the procedures has yet to be established and revision are still needed to ensure that the RE10 Chl-a data is comparable to the CyAN CIcyano data. Preliminary analyses indicated general lack of agreement between CIcyano and Chl-a data, and the majority of the disagreement is a combination of differences in sensitivity of the algorithms (Lie et al., *in* prep) as well as a reflection of non-cyanobacterial blooms as RE10 Chl-a data also casts a wider net and includes algae and plants.

9. APPENDIX B.

Table 9.1: List of 238 waterbodies in the final CA custom shapefile for remote sensing data processing. Regional Board is populated with the regional board number and Water Project is populated with if the waterbody is part a the state (indicated with an "S") or federal (indicated with an "F") water project. "-" indicates no COMID, lake name or water project information available.

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
77	Copco Reservoir (Historical)	361267	Suitable	Reservoir	1	-
79	Iron Gate Reservoir (Historical)	361273	Suitable	Reservoir	1	-
90	Freshwater Lagoon	8315523	Limited	Tidal	1	-
91	Stone Lagoon	8315525	Suitable	Tidal	1	-
92	Big Lagoon	8315527	Suitable	Tidal	1	-
267	Indian Tom Lake	2554839	Suitable	Lake	1	-
295	Tule Lake	2554943	Suitable	Agricultural	1	-
530	Lake Pillsbury	8307970	Suitable	Lake	1	-
536	Lake Sonoma	8271433	Suitable	Reservoir	1	-
549	Deadhorse Flat Reservoir	2556873	Limited	Reservoir	1	-
553	Reservoir N	2556901	Limited	Reservoir	1	-
619	Trinity Lake	8245358	Suitable	Reservoir	1	F
990	Clear Lake Reservoir	2554957	Suitable	Reservoir	1	-
1011	Lake Shastina	3917046	Suitable	Reservoir	1	-
1019	Lake Earl	22226558	Suitable	Tidal	1	-
1025	Lake Mendocino	8306696	Suitable	Reservoir	1	-
1087	Silver Lake	14981200	Suitable	Reservoir	5	-
1168	Goose Lake	120052284	Suitable	Lake	5	-
1186	Dorris Brothers Reservoir	7923473	Suitable	Reservoir	5	-
1192	Big Sage Reservoir	7923491	Suitable	Reservoir	5	-
1211	Dorris Reservoir	7926953	Suitable	Reservoir	5	-
1242	Payne Reservoir	7927031	Limited	Reservoir	5	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
1257	Delta Lake	7927093	Suitable	Lake	5	-
1260	West Valley Reservoir	7927103	Suitable	Reservoir	5	-
1336	Snag Lake	7951690	Limited	Lake	5	-
1342	Medicine Lake	7946776	Limited	Lake	5	-
1352	Horr Pond Big Lake	7947624	Suitable	Lake	5	-
1365	Lake Britton	7947664	Suitable	Reservoir	5	-
1386	Thermalito Forebay	7968005	Suitable	Reservoir	5	S
1417	Thermalito Afterbay	7968933	Suitable	Agricultural	5	S
1462	Lake Oroville	12076080	Suitable	Lake	5	S
1620	Camp Far West Reservoir	15012277	Suitable	Reservoir	5	-
1700	Black Butte Lake	7989989	Suitable	Lake	5	F
1890	Rollins Reservoir	15014315	Limited	Reservoir	5	-
2043	Lower Roberts Reservoir	7924879	Suitable	Reservoir	5	-
2103	Bucks Lake	2775024	Suitable	Lake	5	-
2107	Concow Reservoir	2775046	Limited	Reservoir	5	-
2135	Round Valley Reservoir	8026074	Suitable	Reservoir	5	-
2137	Sly Creek Reservoir	8036781	Suitable	Lake	5	-
2144	Whiskeytown Lake	2781993	Suitable	Reservoir	5	-
2288	East Park Reservoir	7993809	Suitable	Lake	5	-
2297	Indian Valley Reservoir	8005383	Suitable	Lake	5	-
2307	McCreary Lake	8014709	Limited	Reservoir	5	-
2333	Lake Spaulding	8063073	Suitable	Lake	5	-
2359	Shasta Lake	120054085	Suitable	Reservoir	5	F
2418	New Bullards Bar Reservoir	8060079	Suitable	Reservoir	5	-
2464	Thurston Lake	8008511	Suitable	Lake	5	-
2554	Frenchman Lake	8038615	Suitable	Reservoir	5	-
2555	Lake Davis	8038617	Suitable	Lake	5	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
2572	Gold Lake	8038673	Limited	Lake	5	-
2581	Scotts Flat Reservoir	8063155	Suitable	Reservoir	5	-
2599	Jackson Meadows Reservoir	120053436	Limited	Reservoir	5	-
2603	Bowman Lake	8062949	Suitable	Reservoir	5	-
2608	French Lake	8062965	Limited	Reservoir	5	-
2621	Fordyce Lake	8063005	Limited	Reservoir	5	-
2672	Union Valley Reservoir	14981052	Suitable	Reservoir	5	-
2728	Mountain Meadows Reservoir	20295245	Suitable	Reservoir	5	-
2828	Loon Lake	120053833	Suitable	Reservoir	5	-
2925	Moon Lake	7927141	Suitable	Reservoir	5	-
2942	Blue Lake	7927117	Limited	Lake	5	-
2972	Stony Gorge Reservoir	7990259	Suitable	Reservoir	5	-
3005	Juniper Lake	2772452	Limited	Lake	5	-
3024	Clear Lake	8005399	Suitable	Lake	5	-
3062	Butt Valley Reservoir	2772516	Suitable	Reservoir	5	-
3075	Little Grass Valley Reservoir	120053229	Suitable	Reservoir	5	-
3126	Ice House Reservoir	14981136	Suitable	Reservoir	5	-
3154	Renner Lake	120052285	Suitable	Lake	5	-
3159	Bayley Reservoir	7927085	Suitable	Reservoir	5	-
3296	French Meadows Reservoir	14991593	Suitable	Reservoir	5	-
3299	Merle Collins Reservoir	8060125	Suitable	Reservoir	5	-
3300	Hell Hole Reservoir	14991599	Suitable	Reservoir	5	-
3302	Folsom Lake	20194524	Suitable	Reservoir	5	F
3311	Lake Almanor	2772500	Suitable	Reservoir	5	-
3313	Antelope Lake	8026058	Suitable	Reservoir	5	S
3327	Twin Lakes Reservoir	14981180	Limited	Reservoir	5	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
3328	Lake Siskiyou	7964705	Suitable	Reservoir	5	-
3353	Lake Berryessa	130951632	Suitable	Reservoir	5	-
3393	Lake Woollomes	17160996	Suitable	Agricultural	5	-
3440	Pine Flat Lake	22057133	Suitable	Reservoir	5	-
3576	Lake Webb	17168272	Suitable	Reservoir	5	-
3718	Courtright Reservoir	22048371	Suitable	Reservoir	5	-
3802	Bravo Lake	17152378	Suitable	Agricultural	5	-
4259	Wishon Reservoir	22048623	Suitable	Reservoir	5	-
4389	Lake Kaweah	120052624	Suitable	Reservoir	5	-
4416	Castac Lake	14948546	Limited	Lake	5	-
4610	Isabella Lake	22670080	Suitable	Reservoir	5	-
4613	Lake Success	14930103	Suitable	Reservoir	5	-
4651	Jennings Ponds	120050792	Suitable	Agricultural	5	-
4678	Lake Alpine	120051897	Limited	Reservoir	5	-
4686	Beardsley Lake	342083	Suitable	Reservoir	5	-
4723	Donnell Lake	345487	Suitable	Reservoir	5	-
4741	Pinecrest Lake	345581	Suitable	Reservoir	5	-
4742	New Melones Lake	120051899	Suitable	Reservoir	5	F
4749	Woodward Reservoir	2822246	Suitable	Reservoir	5	-
4796	Turlock Lake	2822462	Suitable	Reservoir	5	-
4966	Clifton Court Forebay	1890272	Suitable	Agricultural	5	S
5302	Salt Springs Reservoir	120052626	Suitable	Reservoir	5	-
5392	Upper Blue Lake	17053824	Limited	Reservoir	5	-
5405	Pardee Reservoir	17055346	Suitable	Reservoir	5	-
5453	Huntington Lake	17114527	Suitable	Reservoir	5	-
5487	Florence Lake	17117151	Limited	Reservoir	5	-
5488	Desolation Lake	120052656	Limited	Lake	5	-
5508	San Luis Reservoir	17112371	Suitable	Reservoir	5	S, F (partial)
5768	Thousand Island Lake	16603633	Limited	Lake	5	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
5832	Los Banos Reservoir	17080669	Suitable	Reservoir	5	S
5964	Hetch Hetchy Reservoir	21611075	Suitable	Reservoir	5	-
5969	Yosemite Lake	17065952	Suitable	Reservoir	5	-
6073	New Hogan Lake	17052964	Suitable	Reservoir	5	-
6194	Lower Bear River Reservoir	20194786	Suitable	Reservoir	5	-
6196	Jenkinson Lake	20197596	Suitable	Reservoir	5	-
6246	Camanche Reservoir	17080041	Suitable	Reservoir	5	-
6247	Don Pedro Reservoir	17076029	Suitable	Reservoir	5	-
6319	Cherry Lake	17114471	Suitable	Reservoir	5	-
6386	Mammoth Pool Reservoir	17114481	Suitable	Reservoir	5	-
6390	Bass Lake	17114583	Suitable	Reservoir	5	-
6405	Shaver Lake	120053836	Suitable	Reservoir	5	-
6480	Millerton Lake	21608959	Suitable	Reservoir	5	-
6514	Tenaya Lake	2822404	Limited	Lake	5	-
6571	Modesto Reservoir	19769209	Suitable	Reservoir	5	-
6604	O'Neill Forebay	17117691	Suitable	Reservoir	5	S
6646	Martha Lake	120053883	Limited	Lake	5	-
6689	Lake Thomas A Edison	2818128	Suitable	Reservoir	5	-
6704	Salt Spring Valley Reservoir	17080605	Suitable	Reservoir	5	-
6781	Lake Eleanor	17113675	Suitable	Lake	5	-
6783	Duck Lake	21606665	Limited	Lake	5	-
6915	Lake McClure	19770005	Suitable	Reservoir	5	-
6936	South SF Bay Salt Pond A8	120051850	Suitable	Tidal	2	-
6967	South SF Bay Salt Pond A4	17692747	Suitable	Tidal	2	-
7016	South SF Bay Salt Pond A2E	17692753	Suitable	Tidal	2	-
7017	South SF Bay Sunnyvale WPCP Pond	17693795	Suitable	Tidal	2	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
7051	Anderson Lake	2784475	Suitable	Reservoir	2	-
7194	Mallard Reservoir	1670833	Suitable	Reservoir	2	-
7239	San Pablo Reservoir	1670847	Suitable	Reservoir	2	-
7242	Briones Reservoir	2803341	Suitable	Reservoir	2	-
7255	San Andreas Lake	2806071	Suitable	Reservoir	2	-
7339	San Antonio Reservoir	17692669	Suitable	Reservoir	2	-
7385	Upper Crystal Springs Reservoir	5329245	Suitable	Reservoir	2	-
7390	Abbotts Lagoon	2808483	Limited	Tidal	2	-
7557	Lake del Valle	2809463	Suitable	Reservoir	2	S
7558	Calaveras Reservoir	120053056	Suitable	Reservoir	2	-
7580	Lake Hennessey	17686953	Suitable	Reservoir	2	-
7660	Lake Merced	17609317	Limited	Lake	2	-
7689	Lake Cachuma	17657101	Suitable	Reservoir	3	-
7771	Carrizo Plain complex	17671477	Limited	Lake	3	-
7781	Hernandez Reservoir	120053442	Suitable	Reservoir	3	-
7854	Lake San Antonio	17657079	Suitable	Lake	3	-
7970	Soda Lake	8192379	Limited	Lake	3	-
7981	Lopez Lake	8210941	Suitable	Reservoir	3	-
7992	Lake Nacimiento	22522315	Suitable	Reservoir	3	-
8184	Puddingstone Reservoir	22560266	Limited	Reservoir	4	-
8195	Lake Mathews	22532548	Suitable	Reservoir	8	-
8208	Perris Reservoir	22558520	Suitable	Reservoir	8	S
8217	Big Bear Lake	17586364	Suitable	Reservoir	8	-
8240	Lake Casitas	20332372	Suitable	Reservoir	4	-
8262	Sweetwater Reservoir	20334374	Suitable	Reservoir	9	-
8267	San Vicente Reservoir	20329440	Suitable	Reservoir	9	-
8293	Lake Sutherland	22534902	Suitable	Lake	9	-
8417	Lake Hemet	17568947	Limited	Reservoir	8	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
8442	Pyramid Lake	20342929	Suitable	Reservoir	4	S
8471	Lake Henshaw	17573833	Suitable	Lake	9	-
8482	Lake Piru	20332390	Suitable	Reservoir	4	-
8500	Lower Otay Lake	22545095	Suitable	Reservoir	9	-
8523	Skinner Reservoir	20332350	Suitable	Reservoir	9	-
8531	El Capitan Lake	17568961	Suitable	Reservoir	9	-
8550	Bouquet Reservoir	17568963	Suitable	Reservoir	4	-
8551	Elderberry Forebay	22560294	Suitable	Reservoir	4	-
8553	Santiago Reservoir	120053060	Suitable	Reservoir	8	-
8559	Castaic Lagoon	17568965	Suitable	Reservoir	4	S
8568	Castaic Lake	22532616	Suitable	Reservoir	4	S
8584	Lake Elsinore	20314430	Suitable	Lake	8	-
8727	Upper Alkali Lake	20317978	Suitable	Lake	6	-
8738	Middle Alkali Lake	20310407	Limited	Lake	6	-
8828	Dodge Reservoir	20291765	Limited	Reservoir	6	-
8839	Eagle Lake	22657961	Suitable	Lake	6	-
9078	Silverwood Lake	22657963	Suitable	Reservoir	6	-
9079	Lake Arrowhead	20268205	Suitable	Reservoir	6	-
9120	South Haiwee Reservoir	20287066	Suitable	Reservoir	6	-
9130	Gem Lake	20273063	Limited	Lake	6	-
9157	Convict Lake	20277691	Limited	Lake	6	-
9306	Tinemaha Reservoir	20287024	Suitable	Reservoir	6	-
9330	June Lake	20273047	Suitable	Lake	6	-
9358	Lake Crowley	20286922	Suitable	Reservoir	6	-
9471	Saddlebag Lake	20286972	Limited	Lake	6	-
9473	Grant Lake	20268199	Suitable	Lake	6	-
9483	North Haiwee Reservoir	20286504	Suitable	Reservoir	6	-
9486	Mono Lake	22677686	Suitable	Lake	6	-
9501	Lake Palmdale	22597891	Suitable	Reservoir	6	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
9641	Ramer Lake	948100002	Limited	Agricultural	7	-
9650	Salton Sea	0	Suitable	Lake	7	-
10002	Martinez Lake	0	Suitable	Lake	7	-
11444	Topaz Lake	0	Suitable	Reservoir	6	-
13517	Hensley Lake	0	Suitable	Reservoir	5	-
21179	Cascade Lake	0	Limited	Lake	6	-
22941	Los Vaqueros Reservoir	0	Suitable	Reservoir	5	-
24974	Funks Reservoir	0	Suitable	Reservoir	5	-
31370	McLaughlin Mine Tailings Pond	120052758	Limited	Reservoir	5	-
32351	Lake Havasu	8914219	Suitable	Reservoir	7	-
40276	Bridgeport Reservoir	0	Suitable	Reservoir	6	-
41036	Van Norman Bypass Reservoir	0	Limited	Reservoir	4	-
43951	Bull Slough	0	Suitable	Agricultural	5	-
45950	Boca Reservoir	0	Suitable	Reservoir	6	-
48205	Lower Echo Lake	0	Limited	Lake	6	-
48799	Independence Lake	8932994	Limited	Reservoir	6	-
49938	Donner Lake	0	Suitable	Lake	6	-
50956	Senator Wash Reservoir	0	Suitable	Reservoir	7	-
50974	Allensworth Pond	0	Suitable	Agricultural	5	-
65783	Sand ridge Pond	0	Suitable	Agricultural	5	-
88091	Diamond Valley Lake	0	Suitable	Reservoir	9	-
89130	South SF Bay Salt Pond at Mallard Slough	0	Suitable	Tidal	2	-
89279	H. V. Eastman Lake	0	Suitable	Reservoir	5	-
90649	Twin Lakes (Upper)	0	Limited	Lake	6	-
90730	Twin Lakes (Lower)	0	Limited	Lake	6	-
95569	Webber Lake	8943229	Limited	Lake	6	-
107221	Fallen Leaf Lake	0	Suitable	Lake	6	-

Shapefile ID	Waterbody Name	COMID	Suitability for Trend Analyses	Waterbody Type	Regional Board	Water Project
108050	Applegate Lake	0	Suitable	Lake	1	-
109323	Goose Lake	0	Suitable	Lake	7	-
110290	Copper Basin Reservoir	0	Suitable	Reservoir	7	-
110584	Batiquitos Lagoon	0	Limited	Tidal	9	-
112455	Prosser Creek Reservoir	120053784	Suitable	Lake	6	-
113994	Lake Tahoe	0	Suitable	Lake	6	-
116615	Stampede Reservoir	0	Suitable	Reservoir	6	-
150000	Napa River island slough complex	0	Suitable	Tidal	2	-
150001	Prospect Slough	0	Suitable	Tidal	5	S
150002	Horseshoe Bend	0	Limited	Tidal	5	-
150003	San Joaquin River at Bradford Island	0	Suitable	Tidal	5	S, F (partial)
150004	San Pablo Bay	0	Suitable	Tidal	2	-
150005	Lower South SF Bay	0	Suitable	Tidal	2	-
150006	South SF Bay	0	Suitable	Tidal	2	-
150007	Richardson Bay	0	Suitable	Tidal	2	-
150008	Central SF Bay	0	Suitable	Tidal	2	-
150009	New York Slough	0	Suitable	Tidal	2	-
150010	Broad Slough	0	Suitable	Tidal	5	S
150011	Sacramento River at Sherman Island	0	Suitable	Tidal	5	S, F (partial)
150012	Sherman Lake	0	Suitable	Tidal	5	S
150013	Big Break	0	Suitable	Tidal	5	S
150014	Franks Tract	0	Suitable	Tidal	5	S
150015	Suisun Bay	0	Suitable	Tidal	2	S
150016	Carquinesz Strait	17117147	Suitable	Tidal	2	S