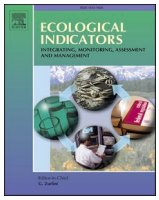
Ecological Indicators 182 (2026) 114492

Contents lists available at ScienceDirect

Ecological Indicators

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



Original Articles

California's approach to standardizing cyanobacterial bloom metrics derived from satellite imagery

Alle A.Y. Lie [a,*], Megan M. Coffer [b,c], Marisa Van Dyke [d], Carly Nilson [d], Jayme Smith [a]

[a] Southern California Coastal Water Research Project Authority, Costa Mesa, CA, USA
[b] Global Science & Technology, Inc., Greenbelt, MD, USA
[c] National Oceanic and Atmospheric Administration, College Park, MD, USA
[d] California State Water Resources Control Board, Sacramento, CA, USA

ARTICLE INFO

Keywords:
FHABs
Cyanobacteria
CI_{cyano}
Annual frequency
Occurrence
Spatial extent
Magnitude

ABSTRACT

Freshwater harmful algal blooms (FHABs) are increasing globally, with implications for public health, ecosystems, and water resource management. Satellite remote sensing complements field-based efforts and has been used to develop large-scale indicators for FHAB monitoring. However, these indicators can be sensitive to processing choices, which may impact regional assessments and management. This study of 83 California waterbodies uses the Cyanobacteria Index (CI_{cyano}) derived from Sentinel-3's Ocean and Land Colour Instrument (OLCI) to systematically evaluate how three common processing decisions (temporal compositing interval, compositing statistics, and pixel masking strategy) affect four widely used FHAB indicators (annual frequency, spatial extent, occurrence, magnitude). Compositing decisions strongly influenced results, with monthly and maximum-value composites yielding the highest bloom metrics, while shorter 7- and 10-day intervals provided more ecologically relevant resolution. Pixel masking strategies involving nearshore regions and ice- and snow-covered areas affected bloom metrics in different ways. Masking more nearshore pixels generally lowered bloom metric values but eliminated some waterbodies from consideration. Ice and snow masking had spatially localized effects, mostly impacting snow-affected regions. Despite differences in absolute values, trend and status analyses using bloom metrics remained largely consistent across masking approaches. Based on the findings, this study recommends a standardized processing workflow for California FHAB monitoring, which may be applicable to other regions with similar management needs: 7-day composites using maximum CI_{cyano} values, and applying both shoreline and ice masking. This workflow balances ecological relevance, computational efficiency, and indicator consistency, supporting the harmonization of bloom monitoring across satellite products, regions, and management programs.

1. Introduction

Freshwater harmful algal blooms (FHABs) are an increasing concern in freshwater systems worldwide, with cyanobacteria posing particular risks due to their potential to produce potent toxins (Huisman et al., 2018; Paerl and Otten, 2013). These blooms can impair drinking water, restrict recreation, and degrade aquatic ecosystems, prompting the need for timely and consistent monitoring. Satellite remote sensing offers a powerful complement to field-based efforts by enabling broad-scale, frequent assessments of surface water conditions that help fill data gaps in regions with many or remote lakes (Papenfus et al., 2020; Wu et al., 2025). These advantages are especially relevant as freshwater management programs seek to detect blooms early, track patterns over time, and prioritize response and management efforts across large landscapes.

As remote sensing becomes increasingly integrated into FHAB monitoring programs, a range of satellite sensors and algorithms have been used, such as the Medium Resolution Imaging Spectrometer (MERIS), the Moderate Resolution Imaging Spectroradiometer (MODIS), and the Ocean and Land Colour Instrument (OLCI) (e.g., Janatian et al., 2025; Laneve et al., 2024; Shahvaran, 2024; Zeng and Binding, 2021). Among these, the Cyanobacteria Index (CI_{cyano}) algorithm (Matthews et al., 2012; Wynne et al., 2010) has become a widely used tool for cyanobacterial bloom detection in the United States, particularly

* Corresponding author.
E-mail address: allel@sccwrp.org (A.A.Y. Lie).

https://doi.org/10.1016/j.ecolind.2025.114492
Received 3 October 2025; Received in revised form 26 November 2025; Accepted 30 November 2025
Available online 6 December 2025

through its implementation in the Cyanobacteria Assessment Network (CyAN).

Cyanobacteria differ optically and ecologically from other algal groups because of their physiological adaptations and accessory pigments, such as phycocyanin and phycoerythrin, which produce distinct spectral features in the visible and near-infrared regions (Bryant, 1982). These spectral differences allow remote-sensing algorithms to distinguish cyanobacteria from total algal biomass typically represented by chlorophyll-*a* (chl-*a*) indices. Phycocyanin-based algorithms directly target the absorption feature near 620 nm, but often exhibit high environmental variability and interference (Simis et al., 2005). In contrast, CI_{cyano} is a shape algorithm that detects an enhanced red-band absorption trough near 681 nm caused by the combined effects of chl-*a* and phycocyanin, thus providing a more robust indicator of cyanobacterial dominance than pigment-specific proxies (Wynne et al., 2010).

CyAN publicly distributes a CI_{cyano} product that is standardized, atmospherically corrected, and georeferenced (Schaeffer et al., 2015). This data product enables the calculation of bloom metrics, including annual frequency (Clark et al., 2017; Coffer et al., 2021a, Coffer et al., 2021b), occurrence (Coffer et al., 2020), spatial extent (Schaeffer et al., 2022; Urquhart et al., 2017), and magnitude (Mishra et al., 2019), which serve as indicators to support assessments and interpretations of bloom status and temporal trends. Despite this standardization at the data production level, there is no widely adopted set of standard processing procedures for how CI_{cyano}, or other remotely sensed water quality data, should be processed post-download to support FHAB assessments. Different agencies and research groups have applied a variety of approaches, often selecting temporal compositing intervals (e.g., 7-day, 10-day, monthly), compositing statistical summaries (e.g., maximum, median, 90th percentile), and masking strategies based on program-specific needs, data availability, or legacy practices (Coffer et al., 2020; McClain et al., 2004; San Francisco Estuary Institute, 2024; Schaeffer et al., 2022; Seegers et al., 2021; Wynne et al., 2022). In addition, nearshore satellite pixels may be affected by both the presence of land at a sub-pixel scale (i.e., a mixed land-water pixel containing both land and water), contamination of reflectance signals from adjacent terrestrial surface ('land-adjacency effect'; Jiang et al., 2023), as well as impacts by bottom reflectance in shallow regions (Zhang et al., 2018). Similarly, ice and snow cover can also alter spectral characteristics in ways that confound algal biomass detection (Wynne et al., 2021).

Earlier versions of the CI_{cyano} data product attempted to correct for nearshore pixels and snow and ice cover, but challenges with false positives remained (Urquhart and Schaeffer, 2020). For example, the extent to which mixed land-water pixels, bottom reflectance, or land-adjacency affects land pixel flagging remains unclear (Jiang et al., 2023; Wang and Qin, 2025; Zhang et al., 2018). Thus, additional masking procedures first defined in Urquhart and Schaeffer (2020) are conventionally performed (e.g., Coffer et al., 2025; Handler et al., 2023; Schaeffer et al., 2022). These additional masking procedures are quality-assurance masking procedures for downloaded Level-3 data aimed to further exclude nearshore pixels and those potentially impacted by snow and ice from downstream analysis. While current versions of CI_{cyano} have greatly improved consideration of potentially confounding conditions, users may elect to still apply additional masking as a conservative approach to reduce the risk of false-positive bloom events.

While compositing and masking choices are expected to impact bloom metric values and influence interpretations of bloom presence, severity, and trends, their effects have not been quantitatively evaluated or documented. Understanding when such processing affects bloom metrics helps determine whether these steps are necessary for specific applications. Since bloom metrics are often used to infer ecological conditions (e.g., an indication of elevated cyanobacterial biomass or bloom persistence), differences in how these metrics are calculated can lead to divergent ecological interpretations. The lack of standardized workflows and limited understanding of their impacts on bloom metrics hinders comparisons across regions, seasons or years, and monitoring programs.

California faces a growing need for consistent, scalable FHAB monitoring and assessment as blooms are increasingly reported across the state's diverse freshwater systems (Smith et al., 2021). With thousands of lakes varying widely in size, accessibility, and climatic context (Franklin and MacDonald, 2024), traditional monitoring approaches are often insufficient to track bloom dynamics across space and time. In response, California recognizes that satellite remote sensing is a cost-effective and complementary approach to field-based assessments of HABs, and has made strategic investments to strengthen its remote sensing program and expand the methods available for routine monitoring (Smith et al., 2021). California's FHAB program thus requires a consistent and transparent remote sensing data processing workflow that reflects its monitoring objectives and the diversity of waterbodies and environmental conditions across the state. Such workflow is vital for supporting public health advisories and informing management decisions.

The goal of this study was to evaluate how common post-download processing choices can influence the calculation of freshwater bloom metrics derived from OLCI CI_{cyano} data. Specifically, the effects of temporal compositing interval, compositing statistics, and pixel masking strategies (nearshore and ice- and snow-affected pixels) on temporal frequency, spatial extent, occurrence, and magnitude were assessed using a six-year CI_{cyano} dataset from 83 of California's large inland waterbodies. While this study was designed to support California's programmatic needs, the workflow is broadly applicable to other regions working to implement or refine remote sensing-based FHAB assessments to ensure consistency in satellite data processing and monitoring.

2. Methods

2.1. Waterbody selection

A total of 83 large inland waterbodies across California were analyzed in this study (Fig. 1). These waterbodies represent the complete set of OLCI-resolvable inland lakes in the state, as defined by Urquhart and Schaeffer (2020) based on the National Hydrography Dataset (NHD). Waterbodies were considered resolvable if their surface area contained at least three 300-m OLCI pixels. The original CyAN shapefile included 88 lakes within or intersecting California, but manual inspections of each polygon revealed some polygons to be segments of the same waterbody. Merging and curation of these segments resulted in a final shapefile with 83 waterbodies in this study (Supplementary Fig. S1).

These 83 waterbodies span a wide range of surface areas (2 – 944 km^2, average of 43 km^2) and encompass a variety of ecosystems (Fig. 1). The area of each OLCI pixel is 0.09 km^2, so the smallest waterbody with an area of 2 km^2 is big enough to accommodate three complete OLCI pixels. The majority are located in central California and fall within or along the Sierra Nevada ($n = 25$) or Central California Foothills and Coastal Mountains ($n = 25$) ecoregions. No waterbodies were located in the Mojave Basin and Range ecoregion, but two (Salton Sea and Lake Havasu) were located in the Sonoran Basin and Range ecoregion, which is characterized by a hot and dry desert landscape in contrast to the cold, mountainous Sierra Nevada ecoregion. Both natural lakes (e.g., Clear Lake and Lake Tahoe) and human-made reservoirs (e.g., Shasta Lake and San Luis Reservoir) were included to reflect a diversity of hydrological and management conditions. Waterbodies that straddle state borders, such as Lake Tahoe and Lake Havasu, were included in the study if they were defined resolvable by Urquhart and Schaeffer (2020).

2.2. Satellite data source

Satellite data were obtained from OLCI onboard the Sentinel-3A and

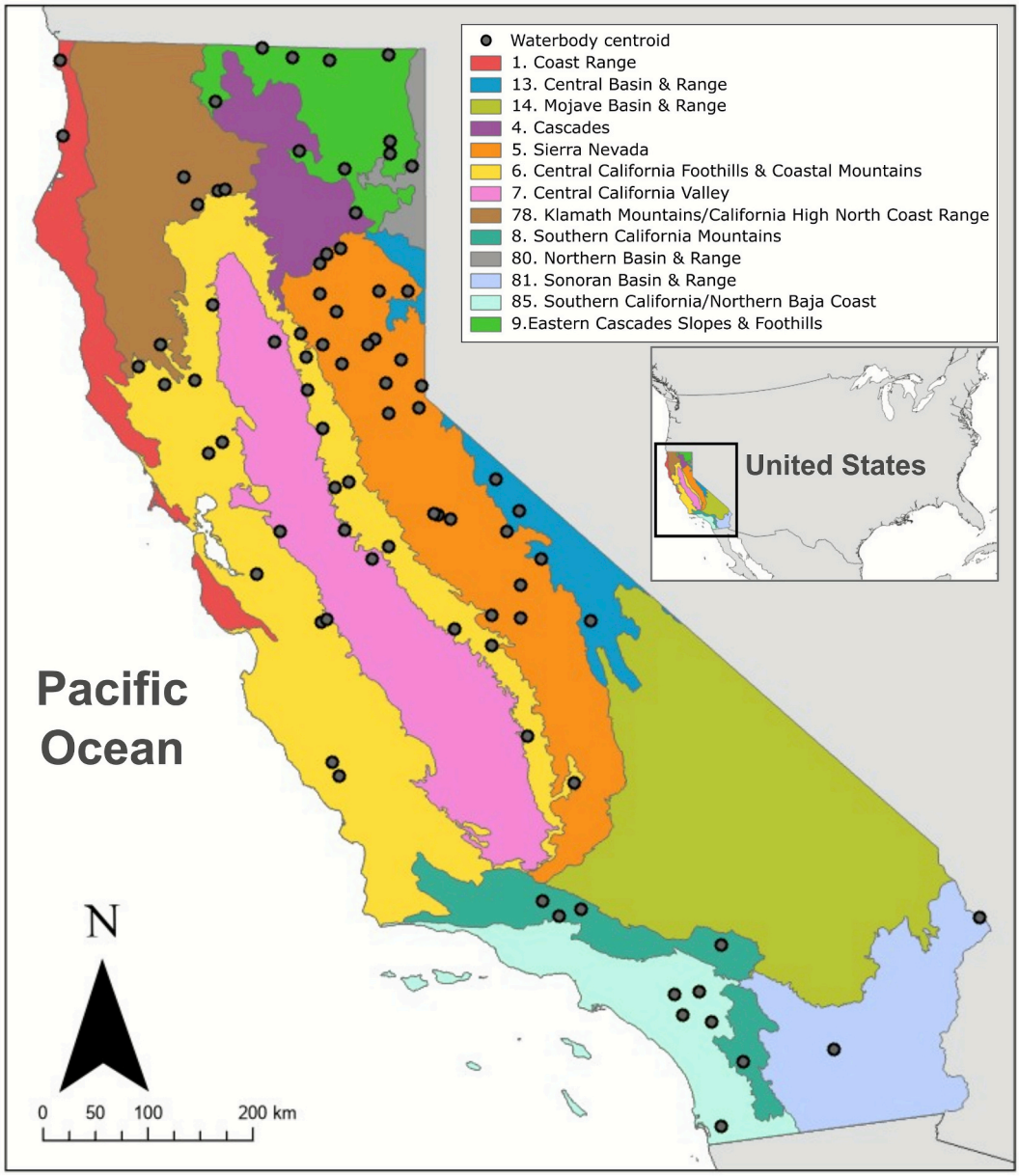


Fig. 1. Map of California showing the centroids of the 83 waterbodies in this study (black points). Colored areas represent the thirteen Level III ecoregions in California. Inset map shows the location of California within the United States of America.

-3B satellites. Sentinel-3A was launched in 2016 followed two years later by Sentinel-3B in 2018, which orbits ~ 140° out of phase of Sentinel-3A. Each satellite is equipped with the OLCI sensor, which collects data in 21 spectral bands spanning visible to near infrared wavelengths (400 – 1020 nm) with 12-bit radiometric resolution, and 300-m spatial resolution. The launch of Sentinel-3B improved OLCI's revisit frequency, from 2–4 days over California to 1–2 days. Hereafter, Sentinel-3A and -3B are referred to collectively as Sentinel-3.

Daily Level-3 processed OLCI data spanning 1 January 2019 to 8 January 2025 were downloaded from the CyAN project website (https://oceancolor.gsfc.nasa.gov/about/projects/cyan/) and used in this study. OLCI imagery data from 2016 to 2018 were omitted to avoid potential bias in assessed changes over time, given inconsistent observational frequency as only Sentinel-3A was operational during that period (Coffer et al., 2025). CyAN data for the contiguous United States (CONUS) are divided into geographical tiles, so tiles 1_2, 1_3, 1_4, 2_2, 2_3, and 2_4 were downloaded to fully cover California and stitched into a single daily mosaic.

The CyAN data product was processed by the National Aeronautics and Space Administration (NASA) Ocean Biology Processing Group (OBPG). Processing steps include atmospheric correction, land flagging using 60-m data from the Shuttle Radar Topography Mission (SRTM), cloud flagging, and calculation of CI_{cyano}. Additional flags are applied by NASA to identify ice and snow-covered pixels using the MERIS Differential Snow Index, and to flag adjacency effects where CI_{cyano} is detected but the Maximum Chlorophyll Index is not (Wynne et al., 2021). CI_{cyano} estimates cyanobacterial biomass using a distinct spectral shape feature centered at 681 nm (Wynne et al., 2010, 2008). The algorithm also includes an exclusion criterion based on the spectral shape at 665 nm (SS(665)), which is sensitive to absorption by the cyanobacterial pigment phycocyanin (Lunetta et al., 2015; Matthews et al., 2012). If $SS(665) < 0$, it indicates low phycocyanin absorption and suggests minimal or no presence of cyanobacteria in the water. CI_{cyano} values were stored as Digital Numbers (DN) in the CyAN product, which can be interpreted using published equations on the CyAN project website.

In this study, the effect of various additional post-download

processing steps were compared, with the aim to develop a workflow that improved the consistency and specificity of bloom detection across California's diverse waterbodies. These included compositing decisions, and more conservative masking of nearshore and masking of ice and snow-pixels using external data from the National Snow and Ice Data Center (NSIDC) as proposed by Urquhart and Schaeffer (2020). These masking refinements were implemented to further reduce the risk of false positive bloom detections, particularly in optically complex nearshore zones and colder seasons or in high-elevation regions.

2.3. Compositing temporal interval and statistical summary method

Daily California mosaics of CyAN data (Version 6) were temporally composited across three timeframes (7-day (7D), 10-day (10D), and monthly) to evaluate the effects of different temporal intervals that reflect operational and ecological timescales used in monitoring programs (e.g., San Francisco Estuary Institute, 2024). Shorter intervals (e.g., 3-day) were not used because they have a higher risk of data gaps, particularly for small waterbodies that may have all pixels flagged as invalid (e.g., due to cloud coverage or sunglint) for a given composite. The start date for the 7D composites was 6 January 2019 to align with the dates of the existing 7D composites offered by CyAN. The 10D and monthly composites began on 1 January 2019. Temporal composites were generated by summarizing CI_{cyano} at each satellite pixel via its maximum (max), 90th percentile (p90), and median (med) value to assess the influence of statistical methods on compositing outcomes. While all three statistics were generated for all three timeframes, comparisons among compositing methods focused primarily on the 7D interval, given its prominence in the CyAN project (Urquhart and Schaeffer, 2020) and existing monitoring programs (San Francisco Estuary Institute, 2024).

2.4. Pixel masking strategies

2.4.1. Nearshore masking

In this study, masking refers to the identification and exclusion of pixels for subsequent metric calculations. Hereafter, valid pixels are those retained after removal of pixels flagged in the CyAN product (e.g., land, cloud, no data) and after application of nearshore or ice and snow masking depending on the workflow being evaluated. 7D max composites were used for all comparisons evaluating the effects of masking. The Level-3 CyAN CI_{cyano} data already include automated land and adjacency masking, as described in Section 2.2. Therefore, this study focused on evaluating additional nearshore masking methods that users may apply post-download to further reduce potential reflectance contamination and to assess how these masking methods influence bloom-metric outcomes.

Pixels adjacent to land (hereafter referred to as nearshore-pixels) were optionally removed from subsequent analysis using different masking methods. Five nearshore masking methods were tested (Supplementary Fig. S2) and are described in detail in the following paragraph: 1) No masking of nearshore-pixels; 2) Masking of waterbody boundary pixels (shoreline-pixels); 3) Masking of pixels within 150 m of the shoreline (150-m buffer); 4) Masking of nearshore-pixels within 300 m of the shoreline (300-m buffer); and 5) Masking of nearshore-pixels using a custom nearshore masking reference (custom reference).

The shoreline of each waterbody was defined using the polygon boundary from the curated waterbody shapefile (described in Section 2.1). Any OLCI pixel whose footprint intersected this polygon boundary was classified as a shoreline-pixel and excluded under the shoreline-masking method (Supplementary Fig. S2C). For the 150-m and 300-m buffer methods, the original waterbody polygons were eroded inward by the specified distances using the st_buffer function from the R *sf* package, creating new interior polygons representing areas 150 m or 300 m from shore. Pixels completely or partially outside these new buffered polygons were masked from further processing (Supplementary Fig. S2D – E). The 150-m and 300-m distances were chosen as they represent half and the full width of an OLCI pixel, respectively. The custom masking reference was based on the dataset provided by Urquhart and Schaeffer (2020), which were further manually curated for the 83 waterbodies in California to account for modifications to the NHD polygons, recent fluctuations in water levels due to prolonged drought or land use changes, and resulting shoreline shifts. These updates to Urquhart and Schaeffer (2020) were made by referencing satellite imagery from ArcGIS (Map Version 2023; source dates from 2020 – 2022) and historical imagery (2017 – 2023) from Google Earth.

2.4.2. Ice & snow masking

Pixels potentially containing ice and snow (hereafter referred to as ice & snow-pixels) were optionally excluded from subsequent bloom metric calculations using different ice & snow masking methods to evaluate the effect of retaining or excluding these pixels. Conventional optical algorithms may misclassify thin or mixed ice-water surfaces due to high visible reflectance and low near-infrared transmittance, which can lead to false-positive bloom detection (Urquhart & Schaeffer, 2020). To reduce this risk, CyAN-based workflows commonly apply masking using the Northern Hemisphere Snow and Ice Analysis data from NSIDC generated using the Iterative Multisensor Snow and Ice Mapping System. Following Urquhart & Schaeffer (2020), daily NSIDC snow and ice data (Version 1) were aggregated into 7D, 10D, or monthly maximum ice & snow-pixels extent composites corresponding to the temporal intervals of the CI_{cyano} composite used. The analysis used both the original 4-km dataset and a newer 1-km version to assess differences in masking precision (Supplementary Fig. S3). Internal holes within the ice & snow polygons were optionally filled (i.e., pixels in these holes will also be masked) to represent a conservative masking scenario that minimizes underestimation of ice or snow coverage caused by fragmented or partially frozen areas (Urquhart & Schaeffer, 2020).

Five ice & snow pixel masking methods were tested (Supplementary Fig. S4): 1) No ice & snow masking; 2) Masking of ice & snow pixels using 1-km reference data; 3) Masking of ice & snow pixels using 1-km reference data with internal holes masked; 4) Masking of ice & snow pixels using 4-km reference data; 5) Masking of ice & snow-pixels using 4-km reference data with internal holes masked.

2.5. Cyanobacterial bloom metrics

The annual frequency, spatial extent, occurrence, and magnitude of cyanobacterial blooms were calculated for each waterbody following composite generation and masking operations. Each metric characterizes a different aspect of cyanobacterial blooms: annual frequency quantifies how often blooms occur; spatial extent quantifies how much surface area blooms cover for each waterbody for each compositing interval (i.e., 7D, 10D, or monthly); occurrence indicates bloom presence within each composite and can be summarized annually for each waterbody (annual waterbody-scale occurrence) or across all waterbodies in the state for each composite (weekly state-scale occurrence); and magnitude quantifies the average bloom intensity within each waterbody for each composite. Details and equations used to compute each metric are listed in Supplemental Text S1. Other studies have developed and applied conceptually similar bloom metrics to summarize algal blooms, such as intensity, extent (as total bloom area), duration, and severity (intensity × extent) (Binding et al., 2023, 2018; Shahvaran et al., 2025).

Two bloom thresholds were used in this study: bloom detection and bloom alert. Bloom detection refers to the ability of the OLCI sensor and CI_{cyano} algorithm to identify any cyanobacterial presence, with a threshold set at $DN \geq 1$ ($DN = 0$ indicates no detection of cyanobacterial presence). This detection limit is preliminarily estimated to be between 10,000–20,000 cells mL^{-1} (Coffer et al., 2021a). Bloom alert refers to more intense blooms that may warrant alerts and health advisories. The World Health Organization recommends 12 μg chl-*a* L^{-1} as Alert Level 1

for managing cyanobacteria in recreational waterbodies (Chorus and Welker, 2021), which corresponds to a DN value of 132 using the DN to CI_{cyano} conversion and CI_{cyano} to chl-*a* conversion from Seegers et al. (2021). Accordingly, the bloom alert threshold used in this study is DN $\geq$ 132. Annual frequency, spatial extent, and occurrence were calculated separately for each bloom detection and bloom alert threshold, as these metrics require a threshold to define whether a pixel is classified as a bloom. In contrast, magnitude was computed as a single value across all valid pixels, without applying a threshold.

2.6. Workflow overview

Imagery data underwent the same general processing workflow regardless of the composite generation and masking methods used. Raw daily GeoTIFFs from CyAN were mosaicked into a single file for California, and composites with different temporal intervals and statistical summaries were generated. Pixels identified as 'land' (DN = 254) via SRTM (see Section 2.2.) and 'no data' (DN = 255) from CyAN preprocessing were excluded by assigning a value of 'NA' to reduce computational load during composite generation. After compositing, additional user-defined nearshore and ice & snow masking steps were applied depending on the analysis scenario. Bloom metrics were calculated using the composite data after all applicable masking methods were applied.

All imagery data processing (compositing and masking), bloom metric calculations, and statistical analyses were conducted in R (version 4.4.2) using the R Geographic Information System (GIS) packages *sf* (Version 1.0–19; Pebesma, 2018) and *terra* (Version 1.8–42 Hijmans, 2023).

2.7. Summary of comparisons & statistical analysis

A set of five comparisons was conducted to evaluate how differences in processing workflows affected the resulting bloom metrics and the interpretation of bloom status. Each of the first four comparisons varied a single processing step while holding all others constant (Table 1): 1) temporal compositing interval was compared across 7D, 10D, and monthly composites, all using max values for compositing and applying shoreline (i.e., waterbody boundary, see Section 2.4.1) and ice & snow (4-km reference with holes masked) masking; 2) compositing statistic was compared across max, p90, and med, while maintaining 7D composites and applying shoreline- and ice & snow (4-km reference with holes masked) masking; 3) five nearshore masking strategies were compared, while maintaining 7D max composites and applying ice & snow (4-km reference with holes masked) masking; 4) five ice & snow masking strategies were compared, while maintaining 7D max composites and applying shoreline masking; and 5) four combinations of nearshore and ice & snow masking were compared, while maintaining 7D max composites. Specifically, these four combinations were: A) no masking (neither nearshore nor ice & snow masking applied); B) no nearshore masking (nearshore masking omitted, but ice & snow masking applied using the 4-km reference with holes masked); C) no ice & snow masking (shoreline masking applied, but ice & snow masking omitted); and D) Shoreline and ice & snow masking (both shoreline masking and ice & snow masking using the 4-km reference with holes masked applied).

For each of these five comparisons, the non-parametric Kruskal-Wallis test was used to determine whether bloom metrics differed significantly among processing workflows (Kruskal and Wallis, 1952). Non-parametric tests were used for this analysis as CyAN data are not normally distributed (Coffer et al., 2021b). The Kruskal-Wallis test indicates only whether workflows differ from each other, so post hoc comparisons were performed using Dunn's test with *p*-values adjusted using the Bonferroni method at $\alpha = 0.05$ (Dunn, 1964). The Kruskal-Wallis test was performed using base R, and Dunn's test was performed via the *FSA* package (Version 0.9.6; Ogle, 2022). For statistical

Table 1
Summary of comparisons between processing workflows conducted in this study.

Comparison	Temporal compositing interval	Compositing statistic	Nearshore masking	Ice & snow masking
1. Temporal compositing interval	7D	Max	Shoreline	4-km ice & snow data (holes masked)
	10D	Max	Shoreline	4-km ice & snow data (holes masked)
	Monthly	Max	Shoreline	4-km ice & snow data (holes masked)
2. Compositing statistics	7D	Max	Shoreline	4-km ice & snow data (holes masked)
	7D	p90	Shoreline	4-km ice & snow data (holes masked)
	7D	Med	Shoreline	4-km ice & snow data (holes masked)
3. Nearshore masking	7D	Max	No nearshore masking	4-km ice & snow data (holes masked)
	7D	Max	Shoreline	4-km ice & snow data (holes masked)
	7D	Max	150-m buffer	4-km ice & snow data (holes masked)
	7D	Max	300-m buffer	4-km ice & snow data (holes masked)
	7D	Max	Custom reference	4-km ice & snow data (holes masked)
4. Ice & snow masking	7D	Max	Shoreline	No ice & snow masking
	7D	Max	Shoreline	1-km ice & snow data
	7D	Max	Shoreline	1-km ice & snow data (holes masked)
	7D	Max	Shoreline	4-km ice & snow data
	7D	Max	Shoreline	4-km ice & snow data (holes masked)
5. Combined nearshore and ice & snow masking	7D	Max	No nearshore masking	No ice & snow masking
	7D	Max	Shoreline	No ice & snow masking
	7D	Max	No nearshore masking	4-km ice & snow data (holes masked)
	7D	Max	Shoreline	4-km ice & snow data (holes masked)

assessments, the 7D composite metric values for spatial extent and magnitude were summarized into seasonal means (Supplementary Text S1) to align with broad ecological cycles and to control for type I (false positive) errors, as the sample size of weekly data had > 20,000 observations. Such extremely large sample sizes can produce statistically significant results even when differences are ecologically negligible (Lin et al., 2013).

3. Results

3.1. Compositing temporal interval (7D vs. 10D vs. monthly)

Differences in compositing temporal interval led to significant differences in all calculated bloom metrics (Kruskal-Wallis $p < 0.001$; Fig. 2). Monthly composites consistently produced significantly higher annual frequency, spatial extent, annual waterbody-scale occurrence, and magnitude than 7D and 10D composites (Dunn's test, adjusted $p < 0.05$; Supplementary Table S1). 10D composites also produced significantly higher values than 7D composites for bloom metrics at the bloom detection threshold (DN $\geq$ 1) (Dunn's test, adjusted $p < 0.05$), but did not differ for blooms exceeding the alert threshold (DN $\geq$ 132) nor for magnitude (Dunn's test, adjusted $p > 0.05$).

3.2. Compositing statistics (max vs. p90 vs. med)

Differences in statistical methods used to generate 7D composites

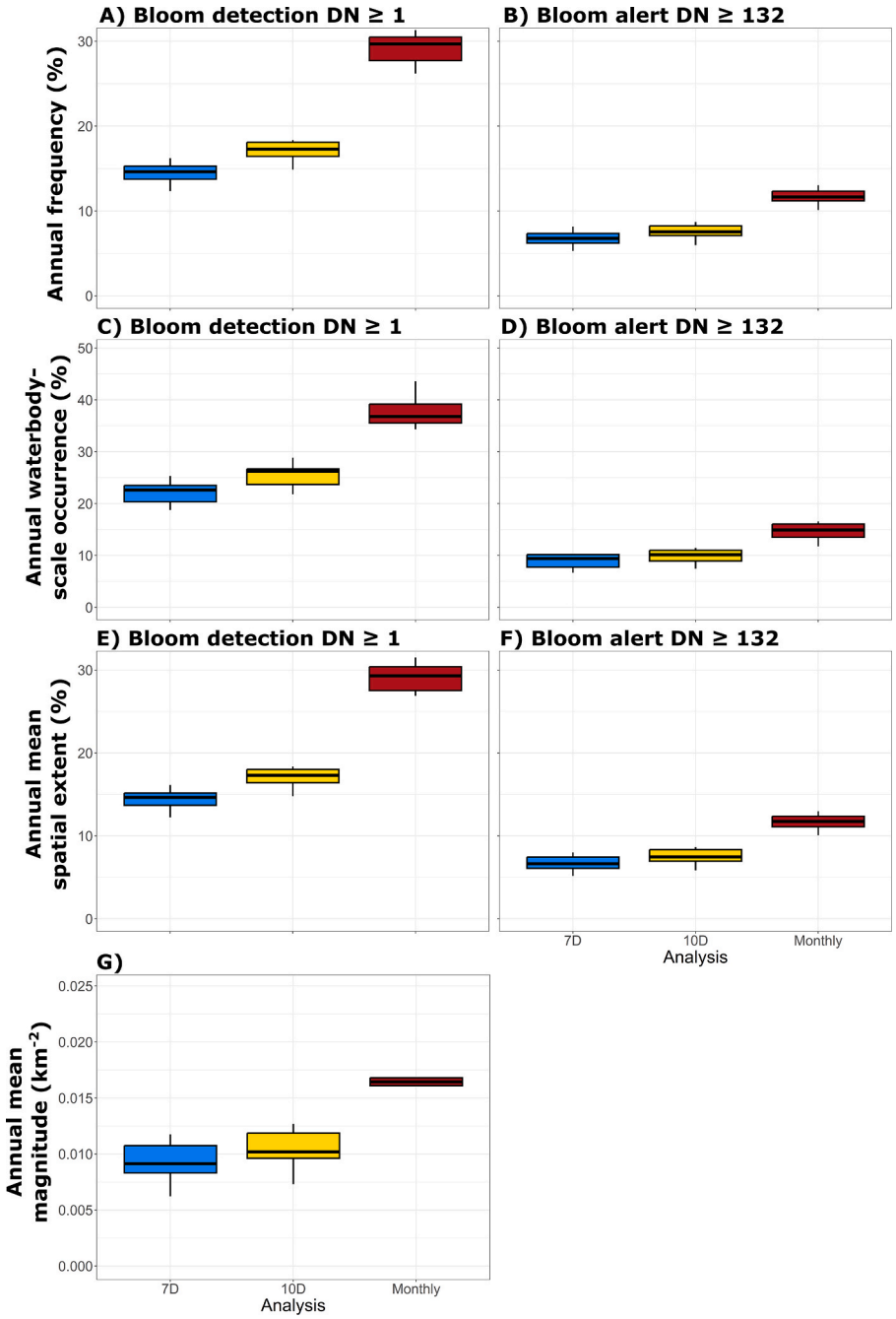


Fig. 2. Boxplots showing the effect of temporal compositing interval (7-day (7D), 10-day (10D), and monthly) on four bloom metrics: annual frequency (A–B), annual waterbody-scale occurrence (C–D), annual mean spatial extent (E–F), and annual mean magnitude (G) across all years. Metrics are shown for both bloom detection (DN $\geq$ 1) and bloom alert (DN $\geq$ 132) thresholds. Outliers are omitted for clarity.

also resulted in significant differences across all bloom metrics (Kruskal-Wallis, $p < 0.001$; Fig. 3, Fig. 4). Med composites consistently produced significantly lower bloom metrics compared to those using max or p90 (Dunn's test, adjusted $p < 0.001$; Supplementary Table S2). There was no significant difference in bloom metrics computed at the bloom detection threshold between max and p90 composites (Dunn's test, adjusted $p > 0.05$), but max composites produced significantly higher annual frequency, spatial extent, annual waterbody-scale occurrence, and weekly state-scale occurrence at the bloom alert threshold, as well as higher magnitude than p90 composites (Dunn's test, adjusted $p < 0.01$).

The effect of compositing statistic varied seasonally, with larger differences in mean weekly values (averaged across years) between methods observed during late spring (week 20) to early fall (week 40), corresponding to the typical cyanobacterial bloom season for California. Larger differences between max and p90 composites were also observed during warmer weeks in the mean weekly state-scale occurrence (Fig. 4B) and spatial extent (Fig. 4D) at the bloom alert threshold, and for magnitude (Fig. 4E). However, week-by-week Kruskal-Wallis tests with Dunn's post hoc comparisons only identified statistically significant differences between max and p90 composites in spatial extent at the bloom alert threshold during colder weeks (weeks < 11 (mid-March) and weeks > 46 (mid-November)). No significant differences between max and p90 composites were found for other weekly bloom metrics.

The aforementioned results focused on comparing different statistical summaries of the 7D composites to isolate the effect of compositing method, but med and p90 composites were also generated for 10D and monthly composites to confirm whether similar trends held across timeframes. Across all temporal intervals, results showed the same general pattern: med composites yielded the lowest bloom metrics, while max composites yielded the highest at the bloom detection threshold (Dunn's test, adjusted $p < 0.01$; Supplementary Fig. S5). For 7D and 10D composites, using max or p90 did not lead to significant differences in bloom metrics, especially at the bloom detection threshold where bloom metrics values were very similar. In contrast, monthly max composites resulted in significantly higher bloom metrics than monthly p90 composites at both bloom thresholds (Kruskal-Wallis and Dunn's test, adjusted $p < 0.001$).

3.3. Nearshore masking

Differences among nearshore masking methods resulted in significant differences in the number of valid pixels used in bloom metric calculations (Kruskal-Wallis, $p < 0.001$). These differences translated to significant differences across bloom metrics using both bloom thresholds (Fig. 5, Fig. 6). All bloom metrics at the bloom alert threshold, as well as annual waterbody-scale occurrence at the bloom detection threshold and magnitude, were significantly different across nearshore-pixel masking methods (Kruskal-Wallis, $p < 0.001$). On the other hand, annual frequency and seasonal mean extent at the bloom detection threshold were statistically similar between masking methods (Kruskal-Wallis, $p > 0.05$). The number of valid pixels that remained after applying a 150-m buffer compared to using a custom reference mask also did not differ significantly (Dunn's test, adjusted $p > 0.05$; Supplementary Table S3), despite spatial differences in the specific nearshore pixels being masked for individual lakes (Supplementary Fig. S2B vs S2D). The six-year average of annual total valid pixels was 1,063,147 and 1,051,778 for the 150-m buffer and custom reference methods, respectively, representing only a ~ 1% difference statewide. Consequently, bloom metric values for both detection and alert thresholds were statistically similar between these two masking methods (Dunn's test, adjusted $p > 0.05$; Supplementary Table S3).

The conservative removal of nearshore-pixels using a 300-m buffer masked a substantial number of pixels (Fig. 5A), removing all potentially resolvable pixels from 14 of the 83 waterbodies in the inventory. This method also resulted in significantly lower values across all bloom metrics and bloom thresholds except annual frequency at the bloom detection threshold compared to all nearshore masking methods, and at the bloom alert threshold when compared to using a 150-m buffer (Dunn's test, adjusted $p < 0.001$; Supplementary Table S3). While the mean weekly spatial extent at the bloom detection threshold (Fig. 6C) and magnitude (Fig. 6E) were higher during warmer weeks (weeks 20 – 40, mid-May through early October) using the 300-m buffer, median values from the same time periods were consistently lower for this masking method across all metrics. This lower central tendency is consistent with the results of the Kruskal-Wallis and Dunn's tests, which rely on rank comparisons and are more closely aligned with medians than means.

Across waterbodies and weeks, the tendency for spatial extent at the bloom alert threshold and magnitude from the 300-m buffer method to

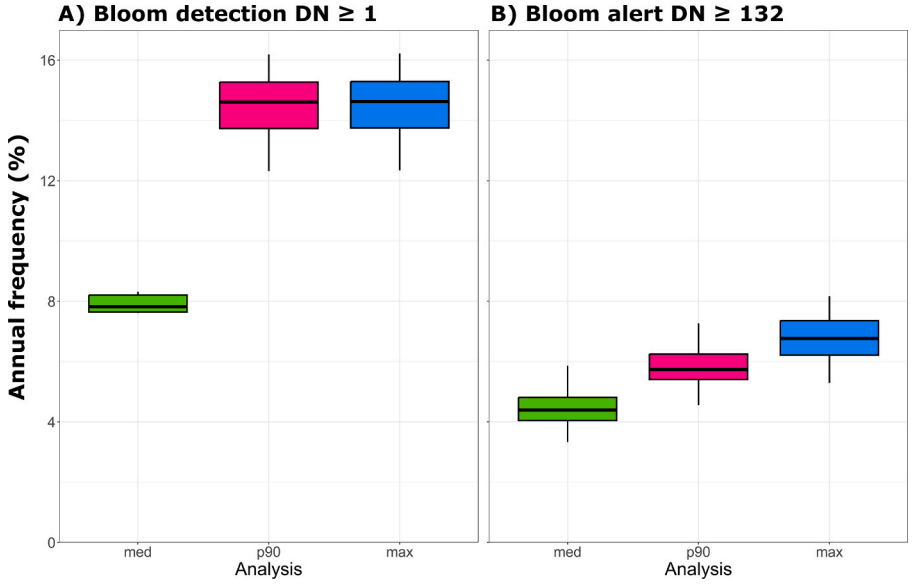


Fig. 3. Boxplots showing the effect of compositing statistics (median (med), 90th percentile (p90), and maximum (max)) on annual frequency across all years for both bloom detection (DN ≥ 1; A) and bloom alert (DN ≥ 132; B) thresholds. Outliers are omitted for clarity.

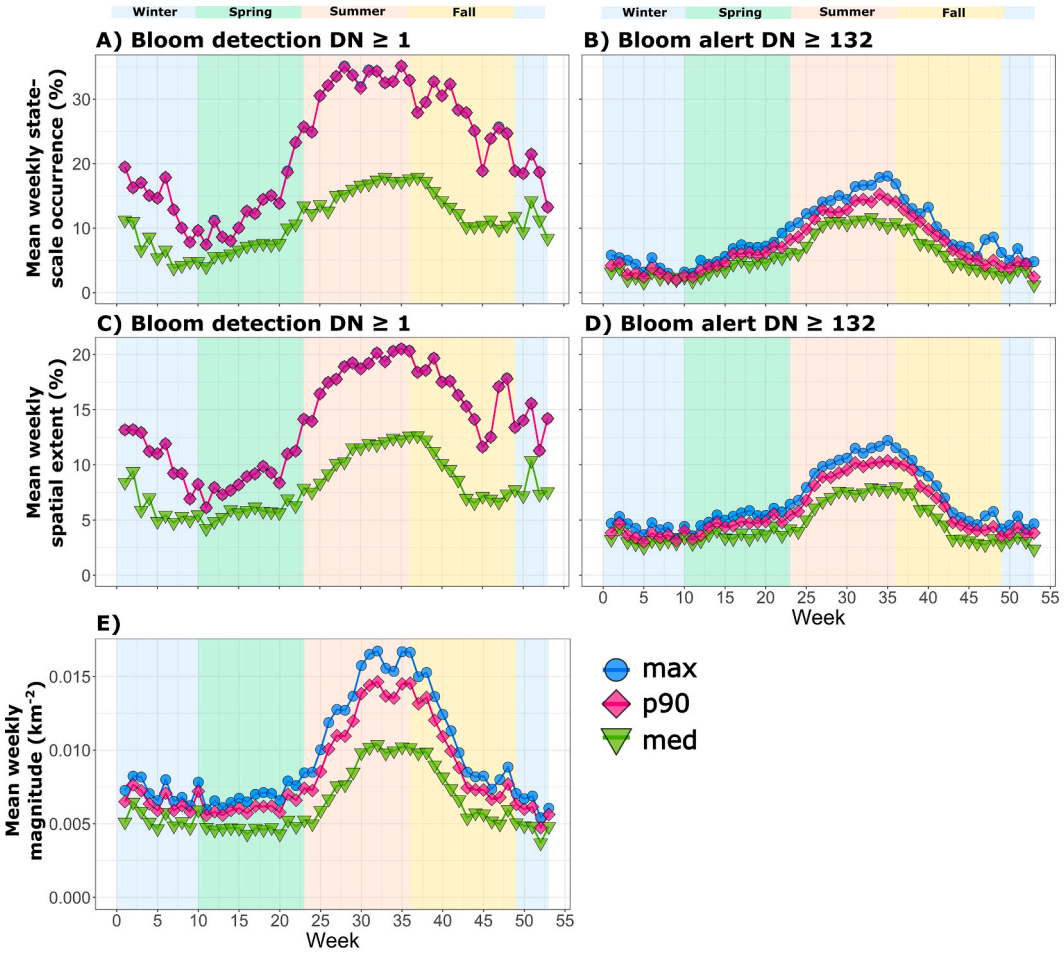


Fig. 4. Mean weekly bloom metrics across all years and waterbodies, showing the effect of compositing statistic (median (med), 90th percentile (p90), and maximum (max)) on the state-scale occurrence (A–B), spatial extent (C–D), and magnitude (E) for both bloom detection (DN $\geq$ 1) and bloom alert (DN $\geq$ 132) thresholds. Background shading indicates seasons for the Northern Hemisphere: winter (late Nov – early Mar, blue), spring (Mar – early Jun, green), summer (Jun – early Sep, red), fall (Sep – Nov, gold). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

exceed values from no nearshore masking was strongly dependent on waterbody size. Logistic regression on log-transformed waterbody resolvable pixel count showed that the odds of the 300-m buffer method yielding higher values than no nearshore masking declined with pixel count (spatial extent odds ratio = 0.82 per unit increase in log(pixels), $p < 0.001$; magnitude odds ratio = 0.70, $p < 0.001$; where an odds ratio < 1 indicates decreasing odds). These results indicate that higher spatial extent at the bloom alert threshold and magnitude using the 300-m buffer method (compared to no nearshore masking) primarily occurred in small waterbodies.

In addition, comparisons of bloom metrics across nearshore masking methods for individual waterbodies and years showed that the effects of nearshore masking varied by waterbody and week (Supplementary Fig. S6). For example, the 300-m buffer method produced either the highest or lowest spatial extent at the bloom detection threshold depending on the waterbody and time of year (Supplementary Fig. S6 top panels). Comparisons of points relative to the 10% occurrence threshold (dashed red line) illustrated instances where no nearshore masking resulted in occurrence while the 300-m buffer method did not at the bloom detection threshold (e.g., Lake Elsinore, week 43), and vice versa (e.g., Clear Lake Reservoir, week 17). Spatial extent was generally lowest under the 300-m buffer method at the bloom alert threshold (middle panels). Values for magnitude varied depending on waterbody and week. For instance, applying a 300-m buffer generally led to the lowest magnitude for Clear Lake in 2024, but the method also led to generally the highest magnitude at Fallen Leaf Lake in mid summer to mid fall (Supplementary Fig. S6 bottom panels).

In contrast, omitting nearshore masking (i.e., preserving nearshore pixels) generally resulted in significantly higher bloom metrics compared to workflows that applied nearshore masking, particularly for metrics using the bloom alert threshold (Supplementary Table S3). Bloom metrics at the bloom alert threshold were also generally higher when no nearshore masking was applied compared to masking only shoreline-pixels (as indicated by positive Z-scores from Dunn's tests with no nearshore masking as Group1), but no statistical differences in bloom metrics at either the bloom detection or bloom alert thresholds were found between the two nearshore masking methods (Dunn's test, adjusted $p > 0.05$), despite a significantly higher number of valid pixels when nearshore-pixels were retained (Dunn's test, adjusted $p < 0.001$).

Additionally, notable seasonal patterns in bloom metrics were observed across nearshore masking methods. Significant differences in magnitude and spatial extent at the bloom alert threshold were most frequently observed during colder weeks (weeks < 13 (before late-March) and > 42 (after mid-October)), while differences in the percentage of occurrence at the bloom detection threshold were more common during warmer weeks (weeks 22–41 (between late May – mid October)) (Kruskal-Wallis test, $p < 0.05$). Post hoc Dunn's tests indicated that omitting nearshore masking often resulted in significantly higher magnitude and spatial extent at the bloom alert threshold during colder periods (adjusted $p < 0.05$).

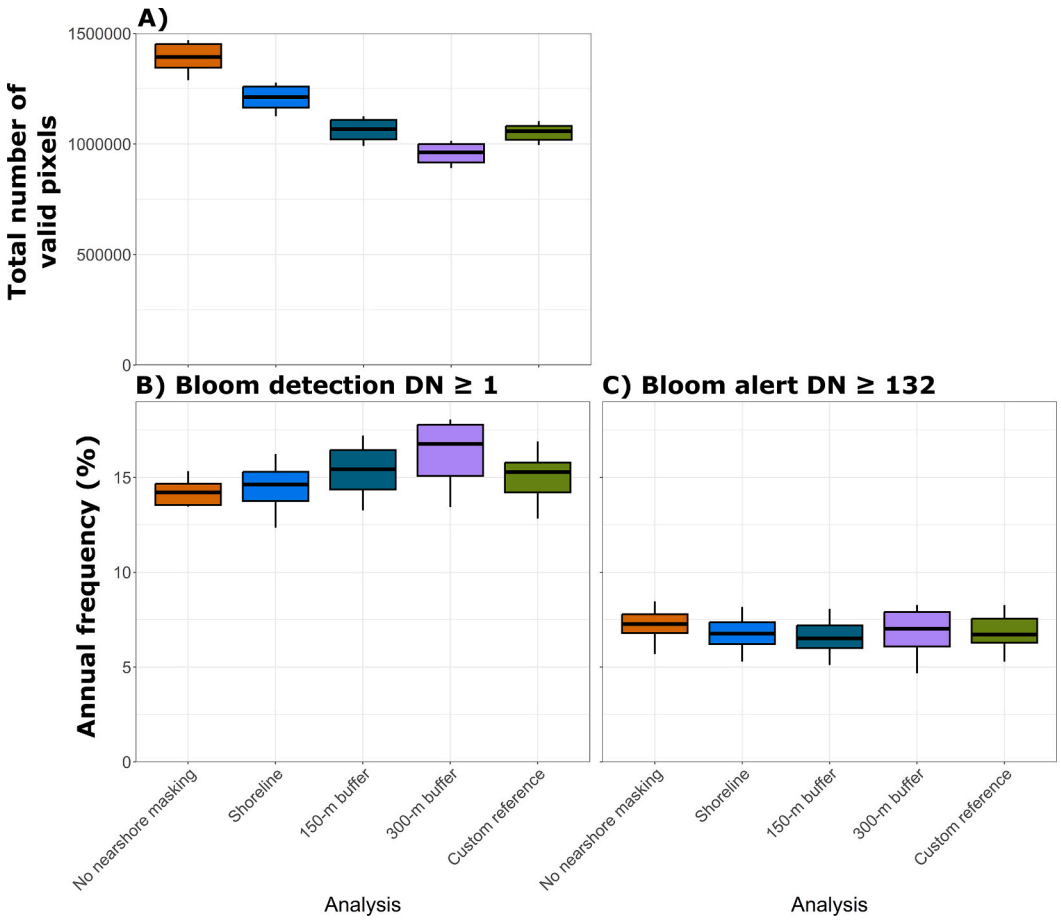


Fig. 5. Boxplots showing the effect of various nearshore masking methods on the total number of valid pixels included in the analysis (A), and on annual frequency for both bloom detection (DN ≥ 1; B) and bloom alert (DN ≥ 132; C) thresholds across all years. Outliers are omitted for clarity.

3.4. Ice & snow masking

Unlike the various nearshore masking methods, masking of potential ice and snow-covered pixels (ice & snow-pixels) did not result in significant differences in the total number of valid pixels (Fig. 7A). Significant differences were only observed for weekly state-scale occurrence at the bloom detection threshold (Kruskal-Wallis test, $p < 0.001$; Fig. 8D), in which the retention of ice & snow-pixels (i.e., no ice-masking) resulted in significantly higher percentage of waterbodies with blooms than applying any type of ice & snow masking (Dunn's test, adjusted $p < 0.01$; Supplementary Table S4). Week-by-week comparisons of the state-scale occurrence ($n = 6$ years per group per week) further showed that these differences were concentrated in colder periods (week < 15 (before mid-April) and > 49 (after mid-December)). Among the masking methods, the 4-km reference with internal holes masked most frequently showed significant differences from no masking (Dunn's test, adjusted $p < 0.05$).

While the various ice & snow masking methods did not lead to significant differences overall, notable regional effects were observed when occurrences at the bloom detection threshold were evaluated on a waterbody-by-waterbody basis. For each waterbody, the number of instances in which an occurrence at the bloom detection threshold was identified when ice & snow-pixels were retained but not under any of the four ice & snow masking methods, was tallied and expressed as a percentage of the total number of weeks (i.e., 7D composites, $n = 313$). Waterbodies showing any difference in occurrences at the bloom detection threshold were primarily located in northern California or within the Sierra Nevada ecoregion (Fig. 9A), with the exception of Big Bear Lake in southern California. Waterbodies with larger disagreement (>18%; gold to red points in Fig. 9A) were all located within the Sierra Nevada ecoregion. The most affected waterbody was Fallen Leaf Lake, just south of Lake Tahoe in Central-Eastern California, which showed a discrepancy in 43% of composites. This indicates that for 43% of weeks, bloom occurrences were identified when ice & snow-pixels were retained, but not when any of the ice & snow masking methods were applied.

The regional pattern was further supported by statistical tests conducted individually for each waterbody (Fig. 9B). Waterbodies with significant Kruskal-Wallis tests ($p < 0.05$) and post hoc Dunn's test indicating a significantly higher annual waterbody-scale occurrence at the bloom detection threshold under the no ice & snow masking method (adjusted $p < 0.05$; red points in Fig. 9B) were exclusively located within the Sierra Nevada ecoregion. Likewise, waterbodies with significant Kruskal-Wallis results but non-significant Dunn's tests also only fell within this ecoregion (orange points in Fig. 9B). In waterbodies where Dunn's test indicated a significant difference, using 4-km reference data with internal holes masked consistently produced a lower annual waterbody-scale occurrence at the bloom detection threshold compared to omitting ice & snow masking.

3.5. Combined nearshore and ice & snow masking

While omitting nearshore masking (vs. shoreline masking) and omitting ice & snow masking (vs. ice & snow masking using the 4-km reference with holes masked) alone generally did not lead to significant differences in derived metrics, the combined retention of both pixel types yielded consistently higher bloom metrics compared to masking them both (positive Z-scores from Dunn's test with no masking as Group

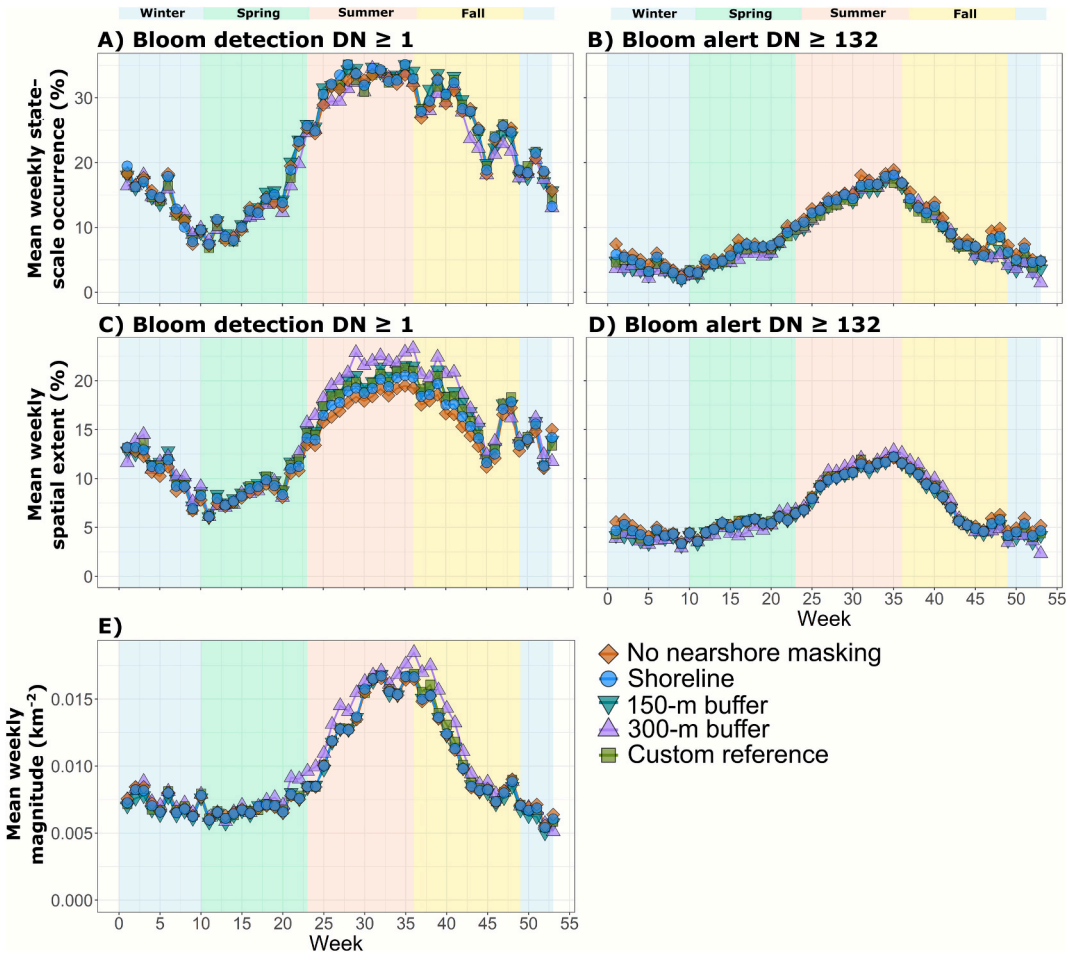


Fig. 6. Mean weekly bloom metrics across all years and waterbodies, showing the effect of various nearshore masking methods on state-scale occurrence (A–B), spatial extent (C–D), and magnitude (E) for both bloom detection (DN ≥ 1) and bloom alert (DN ≥ 132) thresholds. Background shading indicates seasons for the Northern Hemisphere: winter (late Nov – early Mar, blue), spring (Mar – early Jun, green), summer (Jun – early Sep, red), fall (Sep – Nov, gold). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1; Supplementary Table S5), except for annual frequency at the bloom detection threshold (Fig. 10B). However, statistically significant differences were only observed in weekly state-scale occurrence at the bloom detection threshold (Fig. 11) and seasonal mean extent at the bloom alert threshold (Dunn's test, adjusted $p < 0.01$; Supplementary Table S5). Comparisons of the remaining nearshore and ice & snow masking method combinations did not yield additional differences beyond those observed in the individual masking analyses.

Additional consistency analyses were performed to evaluate whether methodological differences could influence downstream trend interpretations by assessing whether different masking methods produced consistent temporal trends. Mann-Kendall trend tests were conducted for each bloom metric (frequency at both thresholds, annual waterbody-scale percentage occurrence at both thresholds, annual mean spatial extent at both thresholds, and annual mean magnitude) across all 83 lakes under the four masking methods included in this comparison (R package *rkt*, version 1.7). Although the Mann-Kendall test has limited statistical power with only six years of data, results showed 98% consistency ($n = 83$ lakes × 7 metrics = 581 tests) in the direction of trends across masking methods (increasing trend: slope > 0, $p < 0.05$; decreasing trend: slope < 0, $p < 0.05$; no trend: $p > 0.05$). To complement this analysis, a ranking-based approach was implemented by classifying lakes into annual quartiles for each metric and assessing agreement in quartile assignments across the masking methods. This analysis also demonstrated high consistency (78%; $n = 83$ lakes × 6 years × 7 metrics = 3,486 comparisons), indicating that the relative ranking of waterbodies was largely robust to processing method. Together, these results demonstrate that while absolute bloom-metric values varied among processing workflows, temporal trends and relative rankings remained consistent, indicating that processing differences did not meaningfully alter trend outcomes.

4. Discussion

This study evaluated how common data processing decisions, specifically temporal compositing intervals, compositing statistics, and pixel masking strategies, influenced the calculation of bloom metrics derived from Sentinel-3 OLCI CI_{cyano} data. These decisions significantly impacted annual frequency, spatial extent, occurrence, and magnitude, with certain processing workflows producing consistently elevated metric values across California's diverse landscape while others had regionally specific effects. Based on these findings, a standardized processing workflow was developed to support California's FHAB monitoring and assessment needs. The recommended workflow uses 7D max composites, shoreline masking, and ice & snow masking based on a 4-km reference with internal holes masked. This workflow is discussed in detail below, and balances ecological relevance, computational efficiency, and data consistency, in order to support reproducible, management-relevant assessments of cyanobacterial bloom patterns.

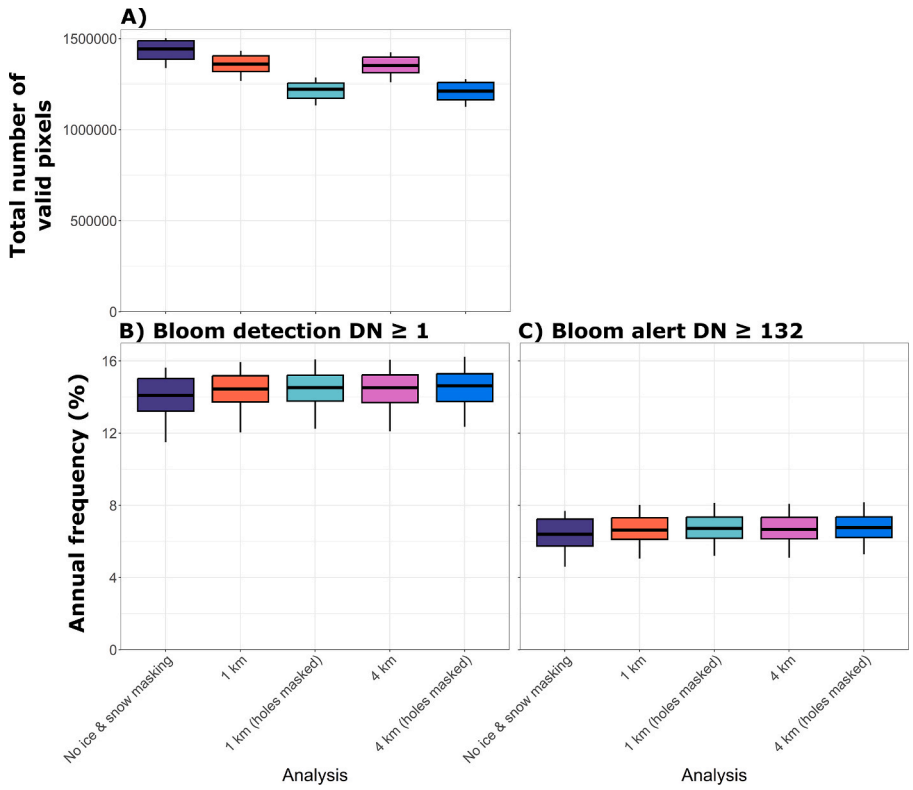


Fig. 7. Boxplots comparing the effect of various ice & snow masking methods on the total number of valid pixels included in the analysis (A), and on annual frequency for both bloom detection (DN ≥ 1; B) and bloom alert (DN ≥ 132; C) across all years. Outliers are omitted for clarity.

4.1. Compositing temporal interval

Temporal compositing interval was one of the strongest drivers of bloom metric variation. This study evaluated 7D, 10D, and monthly composites, which have all been used in previous CI_{cyano}-based FHAB assessments (Coffer et al., 2020; San Francisco Estuary Institute, 2024; Schaeffer et al., 2022; Seegers et al., 2021; Urquhart et al., 2017; Wynne et al., 2022). Monthly composites aggregate more daily observations and can capture multiple bloom peaks within a single interval. This broader temporal window increases the likelihood of retaining extreme pixel values, thereby inflating metrics such as magnitude relative to shorter intervals.

Cyanobacterial blooms can form and dissipate within days to a few weeks (Davis et al., 2009; Oliver et al., 2012; Ranjbar et al., 2021), making monthly aggregation too coarse to capture short-lived or sequential bloom events, thereby reducing temporal resolution and ecological specificity. Similar effects of compositing have been observed in other environmental applications, where coarse temporal or spatial aggregation can obscure short-lived events and lead to under- or over-estimation of derived metrics (Ballester et al., 2024; Bayabil et al., 2019; Gao et al., 2024). Chl-*a* concentrations, in particular, have been shown to vary significantly on daily and weekly scales (Elsdon and Connell, 2009). However, the impact of temporal aggregation may be reduced when the aggregated values are subsequently used in modeling or beta-analyses. For example, Lin et al. (2019) found that solar-induced chlorophyll fluorescence models performed comparably when using daily versus monthly aggregated inputs. This result parallels our consistency analyses, in which different masking methods produced similar waterbody-level trends and status classifications despite differences in absolute bloom-metric values among the masking methods.

The previous rationale for using monthly composites stemmed from the potential for gaps in 7D and 10D composites using MERIS data, which had a temporal resolution of 2–3 days (Urquhart et al., 2017). However, near-daily coverage is now possible through the combined use of Sentinel-3A and -3B platforms, reducing the likelihood of data gaps in 7D or 10D composites. Given that 7D and 10D composites require similar manual effort as monthly composites, there is limited justification for monthly aggregation for data products with high temporal resolution like OLCI besides computational power availability and data storage capacity. While 7D composites generally produced more conservative bloom metrics than 10D, the differences between them were modest compared to the larger discrepancies observed between 7D or 10D vs. monthly intervals (Fig. 2). Although the choice between 7D and 10D may depend on user preference, 7D composites are recommended in the California FHABs program workflow for their clearer interpretability, alignment with weekly sampling timeframes commonly used in field-based monitoring programs, and relation to weekly recreational use.

4.2. Compositing statistics

The choice of compositing statistic also significantly influenced bloom metric values, as supported by Coffer et al. (2025). This study compared med, p90, and max summaries, which represent different methods to aggregate pixel values across the temporal composites. Med composites consistently produced the lowest bloom metrics regardless of temporal compositing interval, reflecting their resistance to extreme values. This makes them appropriate for conservative assessments and long-term trend detection, where capturing typical rather than peak conditions may be the priority. In contrast, max composites provide information on extreme scenarios observed in resolvable open-water areas of waterbodies. Although using max values to summarize 7D composites can also inflate bloom metrics, unlike monthly composites, they do so within a narrower, ecologically meaningful timeframe. Furthermore, extremely high biomass levels are often observed when a waterbody is experiencing a bloom event. This makes the compositing choice more appropriate for identifying high-risk scenarios that align with public health and monitoring objectives, particularly in the

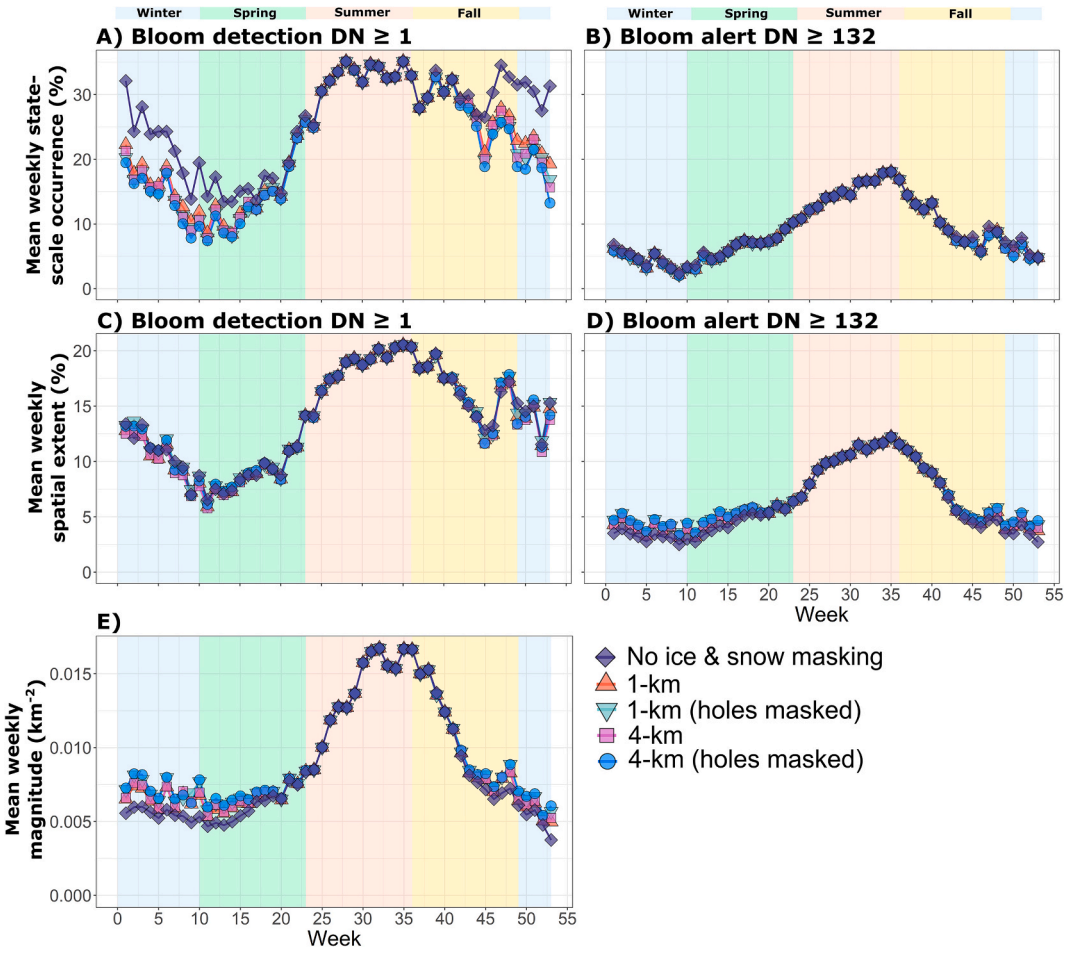


Fig. 8. Mean weekly bloom metrics across all years and waterbodies, showing the effect of various ice & snow masking methods on the state-scale occurrence (A–B), spatial extent (C–D), and magnitude (E) for both bloom detection (DN $\geq$ 1) and bloom alert (DN $\geq$ 132) thresholds. Background shading indicates seasons for the Northern Hemisphere: winter (late Nov – early Mar, blue), spring (Mar – early Jun, green), summer (Jun – early Sep, red), fall (Sep – Nov, gold). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

absence of resolvable shorelines and the lack of field-based cyanotoxin measurements (Ibelings et al., 2021).

Generating composites using p90 instead of max values allows the removal of extreme outliers while still highlighting intense bloom conditions. However, no significant differences between p90 and max composites were found for shorter temporal composites (7D and 10D; Supplementary Fig. S5). Threshold-based metrics (annual frequency, occurrence, and spatial extent) at the bloom detection threshold (DN $\geq$ 1) were particularly similar between p90 and max for 7D and 10D composites. This high similarity reflects the consistency of a pixel being classified as bloom between using either max or p90 due to the relative ease of exceeding the bloom detection threshold. In contrast, monthly composites include many more observations, increasing temporal variability and the likelihood of extreme values, which explains why max and p90 diverge at that timescale. Given that max and p90 composites produced similar bloom metrics at the 7D interval, and that p90 is more computationally intensive, max composites offer a more efficient alternative. Max composites are therefore recommended for programs prioritizing the detection of peak bloom conditions and assessments to inform management decisions involving protection of public health.

4.3. Nearshore masking

Nearshore-pixel masking influenced data availability and bloom metric outcomes, particularly for smaller lakes. While the CyAN Level-3 product already excludes land and "no data" pixels, additional masking of nearshore-pixels is commonly applied to reduce the influence of land-adjacency effects and bottom reflectance (Coffer et al., 2021b; Handler et al., 2023; Seegers et al., 2021). Several masking methods were tested in this study, ranging from no additional nearshore masking to conservative buffers and custom references. Custom references, such as those developed by Urquhart and Schaeffer (2020), offer a customized and often conservative means of identifying nearshore-pixels based on historical shoreline conditions and expert judgment. However, generating these references can be time-intensive and may become outdated as shorelines shift, particularly in regions like California that experience extended periods of drought, where declining water levels expose larger areas of the nearshore zone. Automated removal based on polygon borders also has a similar limitation of relying on potentially outdated shapefiles, but it has the advantage of requiring minimal manual time or effort to implement. While buffer-based methods provide more conservative exclusion, they can reduce data availability, especially for small lakes. For example, applying a 300-m buffer removed all resolvable pixels in 14 of 83 waterbodies, rendering these waterbodies unresolvable. This highlights a central trade-off in bloom metric processing: increasing data quality through masking often comes at the cost of data availability.

In addition to reducing the number of resolvable waterbodies, applying a conservative masking method can also lead to unexpected inflation of bloom metrics for both methodological and ecological reasons. The likelihood of the 300-m buffer method generating higher spatial extent at the bloom alert threshold and magnitude values

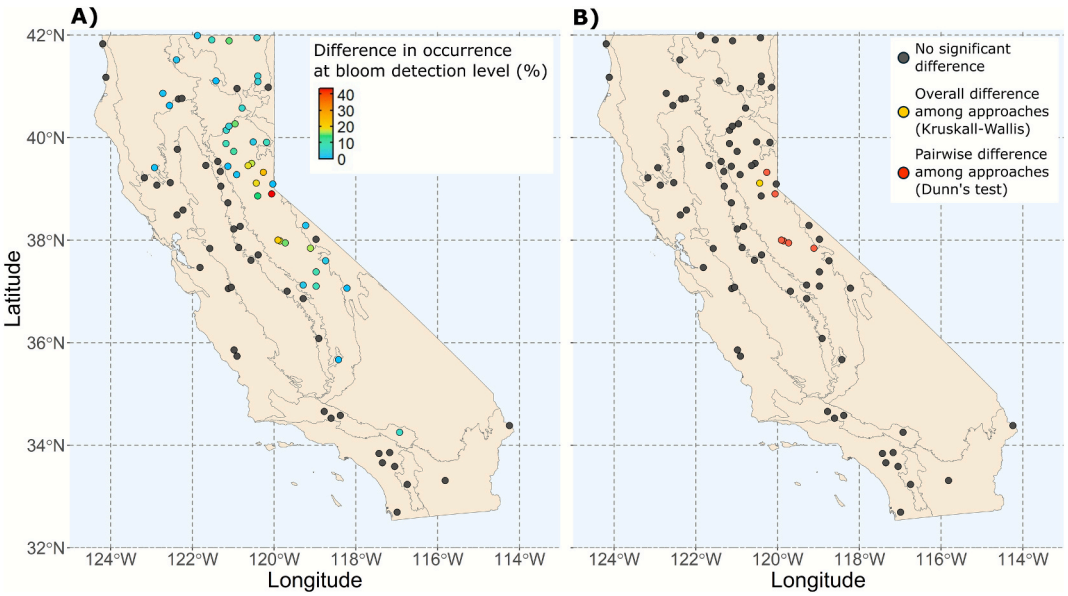


Fig. 9. Maps of California showing the effects of ice & snow masking on occurrence at the bloom detection threshold across waterbodies. Left panel shows the percentage of 7-day composites ($n = 313$) in which occurrence ($\geq$10% of pixels with DN $\geq$ 1) was identified under no ice & snow masking but not when any of the four ice & snow masking methods was applied (A). Black points indicate 0%. The right panel shows results of statistical tests comparing annual waterbody-scale at the bloom detection threshold across ice & snow masking methods for each waterbody, specifically highlighting differences from not applying ice & snow masking ($n = 6$) (B). Yellow points indicate significant Kruskal-Wallis test results ($p < 0.05$) but no Dunn's test significance, and red points indicate significant difference from both Kruskal-Wallis and post-hoc Dunn's tests (adjusted $p < 0.05$). Black points indicate no significant difference between no ice & snow masking and all ice & snow masking methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

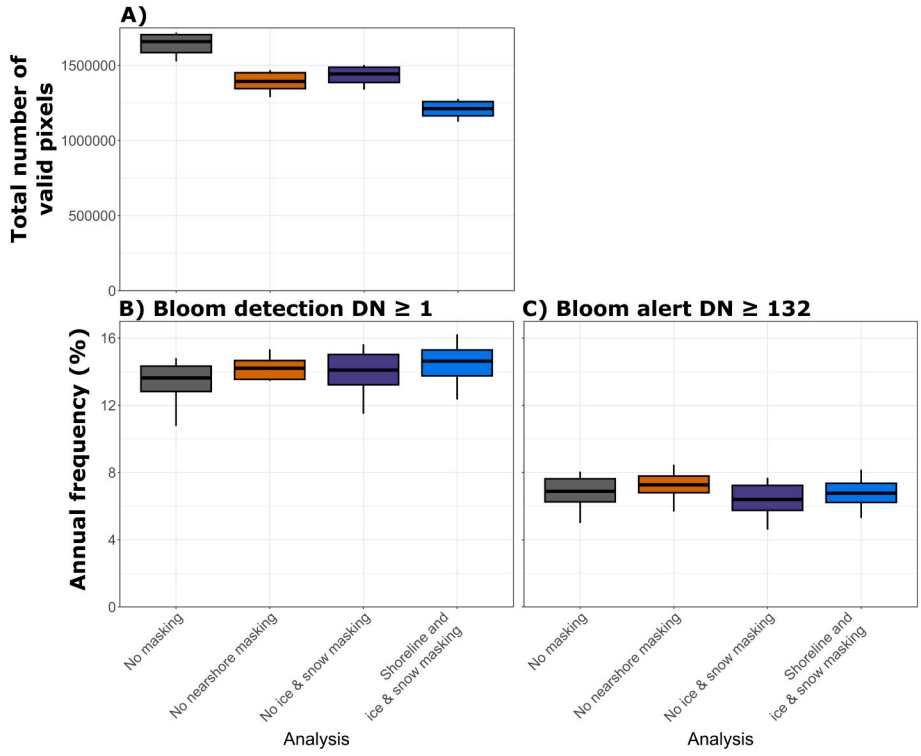


Fig. 10. Boxplots comparing the combined effects of nearshore and ice & snow masking on the total number of valid pixels included in the analysis (A), and on annual frequency for both bloom detection (DN $\geq$ 1; B) and bloom alert (DN $\geq$ 132; C) thresholds across all years. Outliers are omitted for clarity.

(compared to no nearshore masking) was strongly related to waterbody size. In small waterbodies, conservative removal of nearshore pixels disproportionately reduces the denominator (total valid pixels) used to calculate bloom metrics. The remaining interior pixels may also have higher CI_{cyano} values due to land-adjacency effects (Jiang et al., 2023) or because small, shallow systems tend to be more optically and hydrodynamically homogeneous (Holgerson et al., 2022; Mackay, 2011; Vettorazzo et al., 2024). As a result, a few pixels with high CI_{cyano} values can dominate the remaining valid pixel pool, elevating extent or magnitude values. Thus, apparent increases in bloom metrics under

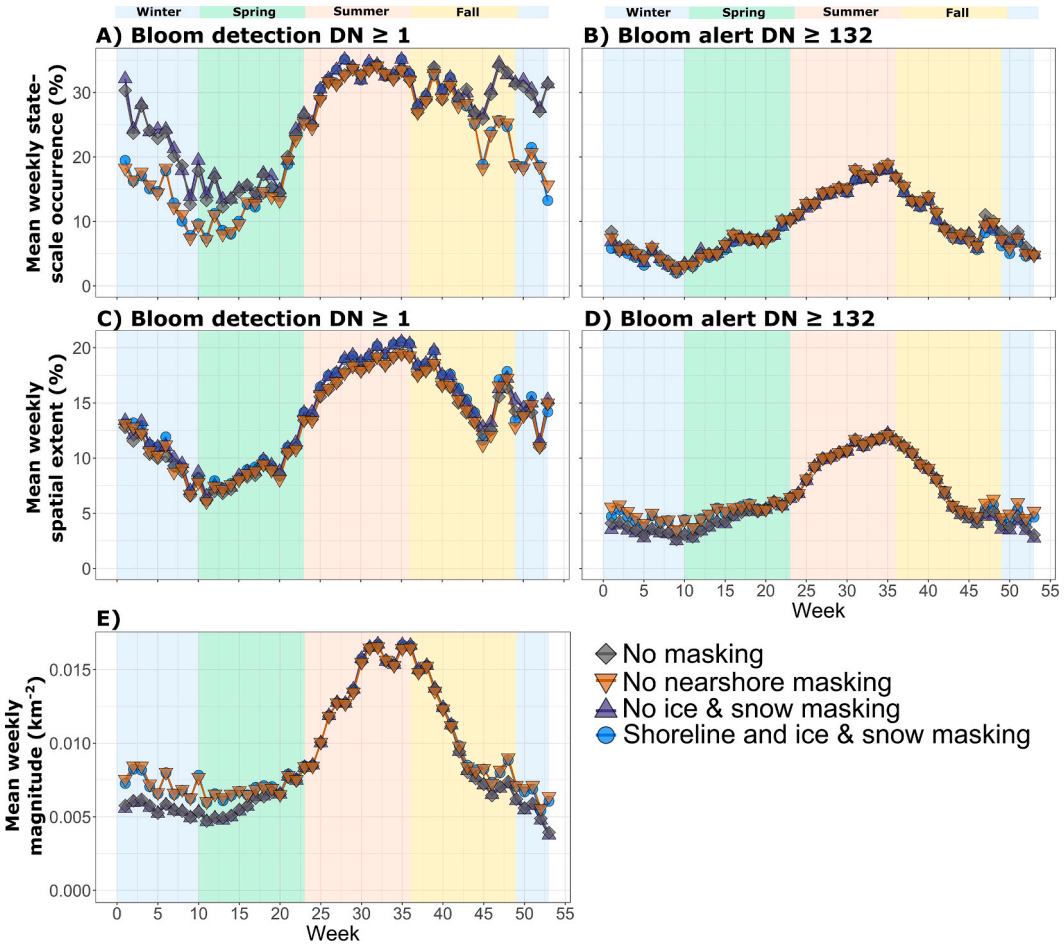


Fig. 11. Mean weekly bloom metrics across all years and waterbodies, comparing the combined effects of nearshore and ice & snow masking on the state-scale occurrence (A–B), spatial extent (C–D), and magnitude (E) for bloom detection (DN ≥ 1) and bloom alert (DN ≥ 132) thresholds. Background shading indicates seasons for the Northern Hemisphere: winter (late Nov – early Mar, blue), spring (Mar – early Jun, green), summer (Jun – early Sep, red), fall (Sep – Nov, gold). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

more conservative masking do not necessarily indicate stronger offshore blooms, but may instead reflect the interplay between pixel removal and the CI_{cyano} values of the remaining pixels. This helps explain the higher weekly mean spatial extent and magnitude values observed under the 300-m buffer method (Fig. 6).

In fact, retaining nearshore-pixels (i.e., no nearshore masking) generally led to significantly higher bloom metric values, indicating that shoreline-pixels tended to have higher DN values. Shorelines can exhibit elevated near-infrared reflectance and distinctive spectral shapes with deeper troughs at 681 nm, potentially inflating CI_{cyano} estimates if SS (665) remains above the CI_{cyano} exclusion threshold (Fig. 4F vs. 4D and Fig. 5H vs. 5E in Jiang et al., 2023). Although cyanobacteria commonly accumulate along shorelines due to hydrodynamics and wind-driven transport (Binding et al., 2023; Wu et al., 2013), these pixels are also prone to contamination of brighter reflectance from terrestrial vegetation and surfaces (Feng and Hu, 2017; Jiang et al., 2023), as well as effects of bottom reflectance (Zhang et al., 2018). Extensive field monitoring would be required to determine whether elevated CI_{cyano} values in shoreline pixels reflect true cyanobacterial blooms or result from erroneous influences. Nevertheless, nearshore areas often hold practical importance for human exposure and monitoring programs, even if they present interpretive challenges in satellite data.

Nearshore sampling may be appropriate in some cases depending on the monitoring objective, such as targeted assessments of accessible nearshore areas that are more relevant to public health concerns (Graham et al., 2008). The shoreline masking method offers a practical compromise by removing only waterbody edge pixels while preserving interior data. This method avoids excessive data loss in small waterbodies while reducing potential false positives from land influences. Even small waterbodies with a single resolvable pixel can still be retained using this method. However, users should be aware of potential data quality concerns in very small waterbodies due to elevated bloom metric values as previously discussed, as well as increased risk of data gaps from all pixels becoming temporarily unresolvable due to cloud cover, drought, or invalid flags obscuring all pixels. This data gap can influence annual summaries (e.g., annual frequency) that depend on how often valid data exist. More sophisticated, data-driven methods for identifying pixels affected by land or nearshore optical signals may further improve masking accuracy (Jiang et al., 2023). However, these methods require access to raw reflectance bands, may be computationally demanding, and are not yet widely validated across diverse waterbodies or specifically evaluated for CI_{cyano}. Their utility in large-scale FHAB assessments remains an important direction for future research. In the meantime, shoreline masking to remove pixels potentially affected by land and nearshore signals is recommended as the preferred method for routine bloom assessments.

4.4. *Ice & snow masking*

In contrast to nearshore masking, the effects of ice & snow masking were spatially localized and significant only for some waterbodies. At Fallen Leaf Lake, for example, omitting ice & snow masking led to a

difference in whether there was bloom occurrence at the bloom detection threshold in 43 % of the 7D max composites, highlighting the potential impact of this processing decision.

Ice & snow-pixels were likely pixels with low-to-moderate DN values ($1 \leq DN < 132$), hence retaining them increased weekly state-scale occurrence at the bloom detection threshold but not at the bloom alert threshold. The lack of an overall significant difference in spatial extent at the bloom detection threshold between applying ice & snow masking and omitting ice & snow masking further indicates that not all ice & snow-pixels exceed the bloom detection threshold. If all ice & snow-pixels had values above this threshold, applying ice & snow masking that removed ice & snow-pixels would reduce spatial extent at the bloom detection threshold. Conversely, if the removed ice & snow-pixels were all actually below the bloom detection threshold (i.e., $DN = 0$), ice & snow masking would decrease the total valid pixel count without affecting the number of bloom-positive pixels, leading to an increase in the calculated spatial extent at the bloom detection threshold. These nuances again emphasize that bloom metrics reflect both pixel intensity (value of DN) and pixel count, so interpretation of masking effects must consider both.

Without extensive field-based observations to compare against, it remains unclear whether DN values in ice & snow-pixels represent true cyanobacteria presence or reflect artifacts from ice & snow. While cyanobacteria are often associated with warm conditions, many taxa persist and even thrive in cold environments (Vincent, 2007; Zakhia et al., 2008) and cyanobacteria have been documented under ice (Bertilsson et al., 2013). This pattern is also supported by observations in California, where blooms have been documented in cooler, high-elevation lakes such as those in the Sierra Nevada mountain range. Limited field monitoring in these regions has confirmed cyanobacterial presence and the risk of cyanotoxins (Brown et al., 2008; Derlet et al., 2009; Streib et al., 2021).

Despite these uncertainties, differences between the retention and removal of ice & snow-pixels were observed in specific waterbodies, supporting the use of consistent ice & snow masking across all waterbodies. While differences amongst ice & snow masking methods were generally minimal, the 4-km ice & snow dataset with internal holes masked emerged as the most conservative option because it excluded the greatest number of ice & snow-pixels and produced the largest differences when compared to omitting ice & snow masking (as indicated by the highest Z-scores among Dunn's pairwise tests; Supplemental Table S4). Additionally, the 4-km product is more practical for many users due to its substantially smaller file size (~13 GB per year) relative to the 1-km dataset (~200 GB per year). For both methodological conservatism and computational efficiency, ice & snow masking using the 4-km dataset with internal holes masked is recommended.

4.5. Seasonal and interannual variability

Seasonal differences between masking methods were evident across many comparisons. Kruskal–Wallis tests comparing weekly state-scale occurrence, spatial extent, and magnitude frequently showed significant differences during colder periods (weeks 1–15 and 40–52). In contrast, larger differences in mean weekly values were often observed during the warmer bloom season (weeks 20–40), and these mean-based differences did not always align with statistical results. For example, the 300-m buffer method showed higher mean spatial extent at the bloom detection threshold in summer (Fig. 6C), despite Kruskal-Wallis indicating that there was no statistically significant difference in the metric and Dunn's Z score indicated the 300-m buffer method generally had lower values (as indicated by positive Z scores with 300-m buffer method being Group 2; Supplementary Table S3). This discrepancy likely reflects the influence of outliers on the mean, as opposed to the rank-based Kruskal-Wallis test, which is more robust to extreme values and better reflects the median. Week-by-week comparisons of individual lakes in 2024 also demonstrated that the relative performance of nearshore masking methods varied by lake and time, with the 300-m buffer sometimes producing the highest or lowest spatial extent at the bloom detection threshold depending on context (Supplementary Fig. S6 top panels). The presence of large mean differences without statistical significance during summer weeks suggests high underlying variability. This variability is expected during the cyanobacterial growing season (Jöhnk et al., 2008; Robson and Hamilton, 2003) and across diverse waterbodies with differing bloom dynamics. Seasonal effects from land-adjacency contamination may also play a role, with stronger influence in summer due to lower sun zenith angles and greener vegetation (Jiang et al., 2023).

This natural variability complicates interpretation of bloom status, particularly in waterbodies with few resolvable pixels. As discussed in section 4.3, masking effects can be particularly pronounced in small lakes, where the removal or retention of a few pixels can substantially influence bloom metrics. Seasonal patterns can amplify this sensitivity. During summer, when blooms can be more spatially heterogeneous (Jöhnk et al., 2008; Verhagen, 1994), compositing and masking decisions can exaggerate short-term fluctuations in spatial extent and magnitude, especially in smaller lakes with limited resolvable area. This also affects whether the waterbody is classified as having a bloom occurrence, complicating public health advisory decisions during this season. In contrast, statistically significant differences were more often detected during winter weeks, which may reflect greater processing sensitivity under less variable ecological and optical (i.e., reduced nearshore reflectance contamination) conditions.

Differences between years due to weather and drought conditions may also affect comparisons across masking methods. Slight interannual variability in comparisons was observed at the individual waterbody level, but overall patterns generally remained the same. For example, omitting nearshore masking generally yielded the highest annual frequency at the bloom alert threshold while using a 300-m buffer zone generally yielded the lowest values, as supported by Kruskal-Wallis and Dunn's tests (Supplemental Table S3). However, exceptions were observed in certain waterbodies and years (Supplementary Fig. S7). For instance, omitting nearshore masking yielded the lowest annual frequency at the bloom alert threshold at Clear Lake Reservoir in 2023 (Supplementary Fig. S7E). Similarly, at Fallen Leaf Lake in 2024, omitting nearshore masking produced the lowest values except when compared to the 300 m buffer method (Supplementary Fig. S7D). Notably, the water years 2020–2022 were classified as 'dry' or 'critically dry' based on the Water Year Hydrologic Classification Index from the California Department of Water Resources (California Department of Water Resources, 2025), while 2023 and 2024 were classified as 'wet' or 'above normal'. These hydrological differences may have influenced shoreline conditions and bloom development, reinforcing the need for a consistent masking strategy to enable more robust comparison across years.

4.6. Bloom threshold considerations

Another important consideration following imagery data processing is the threshold applied for bloom metric calculations. This study evaluated two bloom thresholds: a bloom detection threshold and a bloom alert threshold, and the results indicated that different aspects of the processing workflow affected each threshold differently. For instance, 7D and 10D composites differed significantly in annual frequency only for the bloom detection threshold, while max and p90 composites showed significant differences only for the bloom alert threshold. In addition, nearshore-pixels tended to have values above the bloom alert threshold (as suggested by higher bloom metrics at the bloom alert threshold when nearshore-pixels were retained), whereas ice & snow-pixels were more often above the bloom detection threshold but below the bloom alert threshold (See section 4.4). These interactions between processing steps, pixel types, and bloom thresholds emphasize the need to consider how each element of the workflow affects metric outcomes,

as well as to ensure that threshold selection is guided by the monitoring objective (i.e., whether to detect any cyanobacterial presence or to identify high-risk blooms warranting public health advisories).

The detection threshold reflects the minimum CI_{cyano} signal corresponding to detectable cyanobacterial presence, which is approximately equivalent to 10,000–20,000 cells mL^{-1} (Coffer et al., 2021a). The alert threshold approximates bloom conditions associated with potential human health risk (≈12 μg chl-*a* L^{-1}) according to WHO Alert Level 1 for managing cyanobacteria in recreational waters (Chorus and Welker, 2021). Similarly, other studies have applied thresholds aligned with WHO human health risk categories, such as ≥ 100,000 cells mL^{-1} (DN ≥ 100) for high-risk conditions (Clark et al., 2017; World Health Organization, 1999) or chl-*a* concentration > 10 μg L^{-1} (Binding et al., 2023, 2018; Shahvaran et al., 2025). Together, these thresholds represent ecologically and managerially meaningful benchmarks that support both early detection and public health-oriented assessments of bloom severity.

4.7. Implications for interpretation and management

Differences in bloom metric outputs resulting from processing choices can have meaningful implications for bloom status interpretation and waterbody management. Elevated values from monthly or unmasked composites may overstate bloom prevalence or severity, potentially triggering unnecessary concern or advisories. Conversely, overly conservative methods may underrepresent bloom events, particularly short-lived or shoreline-associated occurrences. As many monitoring programs rely on management criteria (e.g., 10% of spatial extent or annual frequency trends) to identify waterbodies of concern, inconsistencies in metric outcomes may affect lake rankings, bloom classifications, or management actions.

Despite the observed differences in absolute metric values among processing workflows, additional analyses demonstrated that relative temporal and status patterns were largely preserved. Both Mann-Kendall trend and ranking-based consistency analyses showed that temporal trends and relative lake rankings were highly consistent across masking methods (98% and 78% agreement, respectively). For example, Coffer et al. (2021b) ranked waterbodies in California by their annual frequencies from 2019, with Lake Elsinore (Southern California; annual frequency = 100%) among the highest-ranking, Lake Tahoe (Central-Eastern California; annual frequency = 0 %) among the lowest, and Clear Lake (Central California; annual frequency = 20%) in the middle range. The quartile ranks for Lake Elsinore, Lake Tahoe, and Clear Lake were 4 (highest), 1 (lowest), and 3, respectively, based on the quartile ranking analysis in this study, and these rankings were also consistent across all four processing methods tested (see Section 3.5). Overall, these results indicate that while workflow decisions can influence the values of bloom metrics, they have minimal effect on the relative classification of lakes or on the direction of long-term trends.

This study was focused on comparing bloom metrics across processing workflows rather than evaluating CI_{cyano} accuracy against field measurements. Ground-truth validation for CI_{cyano} has been demonstrated in multiple prior studies over the last decade (Coffer et al., 2021a, 2020; Jin et al., 2017; Moradi, 2014; Palmer et al., 2015; Seegers et al., 2021; Whitman et al., 2022). Future efforts should focus on expanding and updating field observations suitable for additional field validation, with an emphasis on California's diverse and high-elevation waterbodies. Specifically, the findings from this study can help guide the design of such future validation efforts. For instance, lakes like Fallen Leaf Lake, where bloom occurrence differs markedly depending on whether ice & snow masking is applied, would be ideal candidates for evaluating cyanobacterial presence under ice- or snow-affected conditions.

While this study focused on freshwater lakes in California, the implications are broadly relevant for remote sensing-based FHAB monitoring in other regions. Optimal workflows should also be tailored to the climatic and geographic context of each monitoring program. In colder regions, consistent application of ice & snow masking remains important for ensuring comparability across years, seasons, and lakes, whereas in temperate or ice-free regions, such masking may be unnecessary. Waterbodies surrounded by dense vegetation or complex shorelines may benefit from nearshore masking to reduce adjacency contamination from terrestrial reflectance. However, bloom metrics like extent and magnitude are normalized by the total number of valid pixels, and may therefore be sensitive to changes in valid-pixel counts introduced by masking. Overly conservative masking can disproportionately reduce data availability and distort bloom metrics in small waterbodies, where the removal of nearshore pixels can strongly influence calculated extent and magnitude values. In addition, regional calibration or additional correction steps may further improve the accuracy of CI_{cyano} values and derived bloom metrics in localized regions or waterbodies, such as those that are optically complex or turbid (Batina and Krtalić, 2024). Workflow decisions should also align with monitoring goals, as approaches optimized for detecting events relevant for public health may require different processing workflows than those designed for long-term ecological trend detection.

Furthermore, workflow choices should consider available computational resources: generating custom composites, especially those based on percentile statistics (e.g., 90th percentile), or incorporating external data sources such as NSIDC for ice & snow masking, can substantially increase data processing and storage requirements. Such considerations are important and conventional in pixel compositing workflows (Adugna et al., 2024; Lv et al., 2023; Meng et al., 2023). For resource-limited programs, an additional advantage of 7D max composites is that CyAN readily offers 7D max composited CI_{cyano} products, minimizing the need for intensive local processing. Ultimately, workflow design should be guided by ecoregional characteristics, management priorities (e.g., public health alerts vs. long-term ecological trends), and practical constraints.

5. Conclusion

This study identified a set of processing recommendations to improve consistency and interpretability in remote sensing-based assessments of FHABs. For California's FHAB program, the recommended workflow includes 7D max composites, shoreline masking to remove nearshore-pixels, and ice & snow masking using a 4-km reference with internal holes masked. This workflow supports weekly decision-making, minimizes contamination from pixels affected by nearshore and ice & snow spectral signals, and preserves data availability across waterbodies with a broad range of shapes, sizes, and climatic settings. While these recommendations were developed to meet the needs of California's statewide monitoring program, the insights are relevant to other regions using remote sensing to track inland water quality, especially given California's diverse hydrological and climatic contexts (Franklin and MacDonald, 2024). The observed sensitivity of bloom metrics to compositing and masking choices underscores the importance of workflow transparency, regardless of geographic region or processing method. These findings may also inform applications using other satellite products, such as chl-*a* estimates from red-edge-based algorithms, and support efforts to harmonize bloom assessment methods across datasets, jurisdictions, and platforms.

CRediT authorship contribution statement

Alle A.Y. Lie: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Megan Coffer:** Writing – review & editing, Supervision, Software, Methodology, Funding acquisition, Conceptualization. **Marisa Van Dyke:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Carly Nilson:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jayme Smith:** Writing – review &

editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Southern California Coastal Water Research Project Informational Technology team for developing Python scripts to support batch downloading of CyAN data, and interns for assistance with testing and validating the data processing scripts. We would also like to acknowledge the Remote Sensing Technical Workgroup and review committee for their thoughtful review of the technical report and for providing valuable feedback that helped improve this study. We are grateful to the two anonymous reviewers for their constructive comments, which helped strengthen the manuscript. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily represent the views or policies of the National Oceanic and Atmospheric Administration (NOAA) or the US Department of Commerce. Mentions of trade names or commercial products do not constitute endorsement or recommendation for use.

Funding sources

This study was supported through funding provided from the State Regional Water Quality Control Board to the Southern California Coastal Water Research Project (agreement number 19-078-270-3). Part of this work was also performed and funded under ST13301CQ0050/1332KP22FNEED0042 from NOAA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2025.114492.

Data availability

The datasets from this study are available in Zenodo, and the code for data processing is available in GitHub. Repository links will be provided upon article acceptance. Custom R scripts used in this analysis are available on GitHub at https://github.com/allel-sccwrp/RemoteSensing. The bloom metric data used in this study are archived on Zenodo at doi: 10.5281/zenodo.17460552. Raw OLCI Level-3 data are publicly available through the CyAN project: https://oceancolor.gsfc.nasa.gov/about/projects/cyan/.

References

Adugna, T., Xu, W., Fan, J., Luo, X., Jia, H., 2024. Multi-temporal pixel-based compositing for cloud removal based on cloud masks developed using classification techniques. Remote Sens. 16. https://doi.org/10.3390/rs16193665.

Ballester, J., van Daalen, K.R., Chen, Z.-Y., Achebak, H., Antó, J.M., Basagaña, X., Robine, J.-M., Herrmann, F.R., Tonne, C., Semenza, J.C., Lowe, R., 2024. The effect of temporal data aggregation to assess the impact of changing temperatures in Europe: an epidemiological modelling study. Lancet Reg. Health – Eur. 36. https://doi.org/10.1016/j.lanepe.2023.100779.

Batina, A., Krtalić, A., 2024. Integrating remote sensing methods for monitoring lake water quality: a comprehensive review. Hydrology 11. https://doi.org/10.3390/hydrology11070092.

Bayabil, H.K., Fares, A., Sharif, H.O., Ghebreyesus, D.T., Moreno, H.A., 2019. Effects of spatial and temporal data aggregation on the performance of the multi-radar multi-sensor system. JAWRA J. Am. Water Resour. Assoc. 55, 1492–1504. https://doi.org/10.1111/1752-1688.12799.

Bertilsson, S., Burgin, A., Carey, C.C., Fey, S.B., Grossart, H.-P., Grubisic, L.M., Jones, I. D., Kirillin, G., Lennon, J.T., Shade, A., Smyth, R.L., 2013. The under-ice microbiome of seasonally frozen lakes. Limnol. Oceanogr. 58, 1998–2012. https://doi.org/10.4319/lo.2013.58.6.1998.

Binding, C.E., Greenberg, T.A., McCullough, G., Watson, S.B., Page, E., 2018. An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. J. Gt. Lakes Res. 44, 436–446. https://doi.org/10.1016/j.jglr.2018.04.001.

Binding, C.E., Zeng, C., Pizzolato, L., Booth, C., Valipour, R., Fong, P., Zastepa, A., Pascoe, T., 2023. Reporting on the status, trends, and drivers of algal blooms on Lake of the Woods using satellite-derived bloom indices (2002–2021). J. Gt. Lakes Res. 49, 32–43. https://doi.org/10.1016/j.jglr.2022.12.007.

Brown, L.R., May, J.T., Hunsaker, C.T., 2008. Species Composition and Habitat Associations of Benthic Algal Assemblages in Headwater Streams of the Sierra Nevada, California. West. North Am. Nat. 68, 194–209. https://doi.org/10.3398/1527-0904(2008)68%255B194:SCAHAO%255D2.0.CO;2.

Bryant, D.A., 1982. Phycoerythrocyanin and Phycoerythrin: Properties and Occurrence in Cyanobacteria. Microbiology. https://doi.org/10.1099/00221287-128-4-835.

California Department of Water Resources, 2025. Water Year Hydrologic Classification Index [WWW Document]. URL https://cdec.water.ca.gov/cgi-progs/iodir/WSIHIST.

Chorus, I., Welker, M., 2021. Toxic Cyanobacteria in Water: a Guide to their Public Health Consequences, monitoring and Management, 2nd ed. World Health Organization, Geneva, Switzerland. https://doi.org/10.1201/9781003081449.

Clark, J.M., Schaeffer, B.A., Darling, J.A., Urquhart, E.A., Johnston, J.M., Ignatius, A.R., Myer, M.H., Loftin, K.A., Werdell, P.J., Stumpf, R.P., 2017. Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. Ecol. Ind. 80, 84–95. https://doi.org/10.1016/j.ecolind.2017.04.046.

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

Coffer, M.M., Schaeffer, B.A., Foreman, K., Porteous, A., Loftin, K.A., Stumpf, R.P., Werdell, P.J., Urquhart, E., Albert, R.J., Darling, J.A., 2021a. Assessing cyanobacterial frequency and abundance at surface waters near drinking water intakes across the United States. Water Res. 201, 117377. https://doi.org/10.1016/j.watres.2021.117377.

Coffer, M.M., Schaeffer, B.A., Salls, W.B., Minucci, J.M., Cronin-Golomb, O., 2025. Recommendations for temporal aggregation of water quality data from multi-platform satellite constellations. Int. J. Remote Sens. https://doi.org/10.1080/01431161.2025.2575515.

Coffer, M.M., Schaeffer, B.A., Salls, W.B., Urquhart, E., Loftin, K.A., Stumpf, R.P., Werdell, P.J., Darling, J.A., 2021b. Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales. Ecol. Ind. 128, 107822. https://doi.org/10.1016/j.ecolind.2021.107822.

Davis, T.W., Berry, D.L., Boyer, G.L., Gobler, C.J., 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. This Issue Contains Spec. Sect. Strains 8, 715–725. https://doi.org/10.1016/j.hal.2009.02.004.

Derlet, R.W., Ger, K.A., Richards, J.R., 2009. Algae in Sierra Nevada Mountain wilderness areas: potential Health Hazards. J. Mt. Med. Ecol. 1, 1–10. https://sierranaturenotes.yosemite.ca.us/Algae_Derlet_20100125.html.

Dunn, O.J., 1964. Multiple comparisons using rank sums. Technometrics 6, 241–252. https://doi.org/10.1080/00401706.1964.10490181.

Elsdon, T., Connell, S., 2009. Spatial and temporal monitoring of coastal water quality: refining the way we consider, gather, and interpret patterns. Aquat. Biol. 5, 157–166. https://doi.org/10.3354/ab00146.

Feng, L., Hu, C., 2017. Land adjacency effects on MODIS Aqua top-of-atmosphere radiance in the shortwave infrared: Statistical assessment and correction. J. Geophys. Res. Oceans 122, 4802–4818. https://doi.org/10.1002/2017JC012874.

Franklin, J., MacDonald, G.M., 2024. Climate change and California sustainability—Challenges and solutions. Proc. Natl. Acad. Sci. 121, e2405458121. https://doi.org/10.1073/pnas.2405458121.

Gao, F., Anderson, M., Houborg, R., 2024. Impacts of spatial and temporal resolution on remotely sensed corn and soybean emergence detection. Remote Sens. 16. https://doi.org/10.3390/rs16224145.

Graham, J.L., Loftin, K.A., Ziegler, A.C., Meyer, M.T., 2008. Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs (No. 2328– 0328). US Geological Survey.

Handler, A.M., Compton, J.E., Hill, R.A., Leibowitz, S.G., Schaeffer, B.A., 2023. Identifying lakes at risk of toxic cyanobacterial blooms using satellite imagery and field surveys across the United States. Sci. Total Environ. 869, 161784. https://doi.org/10.1016/j.scitotenv.2023.161784.

Hijmans, R.J.. terra: Spatial Data Analysis. https://cran.r-project.org/web//packages/terra/terra.pdf.

Holgerson, M.A., Richardson, D.C., Roith, J., Bortolotti, L.E., Finlay, K., Hornbach, D.J., Gurung, K., Ness, A., Andersen, M.R., Bansal, S., Finlay, J.C., Cianci-Gaskill, J.A., Hahn, S., Janke, B.D., McDonald, C., Mesman, J.P., North, R.L., Roberts, C.O., Sweetman, J.N., Webb, J.R., 2022. Classifying Mixing Regimes in Ponds and Shallow Lakes. Water Resour. Res. 58, e2022WR032522. https://doi.org/10.1029/2022WR032522.

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. 16, 471–483. https://doi.org/10.1038/s41579-018-0040-1.

Ibelings, B.W., Kurmayer, R., Azevedo, S.M.F.O., Wood, S.A., Chorus, I., Welker, M., 2021. Understanding the occurrence of cyanobacteria and cyanotoxins. In: Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management. CRC Press, pp. 213–294. https://doi.org/10.1201/9781003081449-4.

Janatian, N., Raudsepp, U., Broomandi, P., Fickas, K., Olli, K., Heimovaara, T., Mannik, A., Uiboupin, R., Pahlevan, N., 2025. A review on remote-sensing-based harmful cyanobacterial bloom monitoring services. Remote Sens. Appl.: Soc. Environ. 37, 101488. https://doi.org/10.1016/j.rsase.2025.101488.

Jiang, D., Scholze, J., Liu, X., Simis, S.G.H., Stelzer, K., Müller, D., Hunter, P., Tyler, A., Spyrakos, E., 2023. A data-driven approach to flag land-affected signals in satellite derived water quality from small lakes. Int. J.Appl. Earth Obs. Geoinformation 117, 103188. https://doi.org/10.1016/j.jag.2023.103188.

Jin, Q., Lyu, H., Shi, L., Miao, S., Wu, Z., Li, Y., Wang, Q., 2017. Developing a two-step method for retrieving cyanobacteria abundance from inland eutrophic lakes using MERIS data. Ecol. Ind. 81, 543–554. https://doi.org/10.1016/j.ecolind.2017.06.027.

Jöhnk, K.D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P.M., Stroom, J.M., 2008. Summer heatwaves promote blooms of harmful cyanobacteria. Glob. Chang. Biol. 14, 495–512. https://doi.org/10.1111/j.1365-2486.2007.01510.x.

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 583–621. https://doi.org/10.2307/2280779.

Laneve, G., Téllez, A., Kallikkattil Kuruvila, A., Bruno, M., Messineo, V., 2024. Eutrophication and HAB occurrence control in lakes of different origins: a multi-source remote sensing detection strategy. Remote Sens. 16. https://doi.org/10.3390/rs16101792.

Lin, M., Lucas, H.C., Shmueli, G., 2013. Research commentary: too big to fail: large samples and the p-value problem. Inf. Syst. Res. 24, 906–917. https://www.jstor.org/stable/24700283.

Lin, X., Chen, B., Zhang, H., Wang, F., Chen, J., Guo, L., Kong, Y., 2019. Effects of the Temporal Aggregation and Meteorological Conditions on the Parameter Robustness of OCO-2 SIF-Based and LUE-Based GPP Models for Croplands. Remote Sens. 11. https://doi.org/10.3390/rs11111328.

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

Lv, X., Persello, C., Zhao, W., Huang, X., Hu, Z., Ming, D., Stein, A., 2023. Pruning for image segmentation: improving computational efficiency for large-scale remote sensing applications. ISPRS J. Photogramm. Remote Sens. 202, 13–29. https://doi.org/10.1016/j.isprsjprs.2023.05.024.

Mackay, E.B., 2011. Heterogeneity in Esthwaite Water, a small, Temperate Lake: Consequences for Phosphorus Budgets. Lancaster University, UK. https://eprints.lancs.ac.uk/id/eprint/133578/1/11003693.pdf.

Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. Remote Sens. Environ. 124, 637–652. https://doi.org/10.1016/j.rse.2012.05.032.

McClain, C.R., Feldman, G.C., Hooker, S.B., 2004. An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. Views Ocean Process. Sea-Viewing Wide Field-View Sens. Seawifs Mission 151, 5–42. https://doi.org/10.1016/j.dsr2.2003.11.001.

Meng, X., Xie, S., Sun, L., Liu, L., Han, Y., 2023. Evaluation of temporal compositing algorithms for annual land cover classification using Landsat time series data. Int. J. Digit. Earth 16, 2574–2598. https://doi.org/10.1080/17538947.2023.2230958.

Mishra, S., Stumpf, R.P., Schaeffer, B.A., Werdell, P.J., Loftin, K.A., Meredith, A., 2019. Measurement of cyanobacterial bloom magnitude using satellite remote sensing. Sci. Rep. 9, 18310. https://doi.org/10.1038/s41598-019-54453-y.

Moradi, M., 2014. Comparison of the efficacy of MODIS and MERIS data for detecting cyanobacterial blooms in the southern Caspian Sea. Mar. Pollut. Bull. 87, 311–322. https://doi.org/10.1016/j.marpolbul.2014.06.053.

Ogle, D.H., 2022. FSA: Simple Fisheries Stock Assessment Methods.

Oliver, R.L., Hamilton, D.P., Brookes, J.D., Ganf, G.G., 2012. Physiology, Blooms and Prediction of Planktonic Cyanobacteria. In: Whitton, B.A. (Ed.), Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Netherlands, Dordrecht, pp. 155–194. https://link.springer.com/book/10.1007/978-94-007-3855-3.

Paerl, H.W., Otten, T.G., 2013. Harmful cyanobacterial blooms: causes, consequences, and controls. Microb. Ecol. 65, 995–1010. https://doi.org/10.1007/s00248-012-0159-y.

Palmer, S.C.J., Odermatt, D., Hunter, P.D., Brockmann, C., Présing, M., Balzter, H., Tóth, V.R., 2015. Satellite remote sensing of phytoplankton phenology in Lake Balaton using 10years of MERIS observations. Remote Sens. Environ. 158, 441–452. https://doi.org/10.1016/j.rse.2014.11.021.

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

Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. R J 10, 439–446. https://journal.r-project.org/articles/RJ-2018-009/.

Ranjbar, M.H., Hamilton, D.P., Etemad-Shahidi, A., Helfer, F., 2021. Individual-based modelling of cyanobacteria blooms: physical and physiological processes. Sci. Total Environ. 792, 148418. https://doi.org/10.1016/j.scitotenv.2021.148418.

Robson, B.J., Hamilton, D.P., 2003. Summer flow event induces a cyanobacterial bloom in a seasonal Western Australian estuary. Mar. Freshw. Res. 54, 139–151. https://doi.org/10.1071/MF02090.

San Francisco Estuary Institute. FHAB Satellite Monitoring Tool. https://fhab.sfei.org/.

Schaeffer, B., Loftin, K., Stumpf, R., Werdell, P., 2015. Agencies Collaborate, Develop a Cyanobacteria Assessment Network. Eos 96. https://doi.org/10.1029/2015EO038809.

Schaeffer, B.A., Urquhart, E., Coffer, M., Salls, W., Stumpf, R.P., Loftin, K.A., Jeremy Werdell, P., 2022. Satellites quantify the spatial extent of cyanobacterial blooms across the United States at multiple scales. Ecol. Ind. 140, 108990. https://doi.org/10.1016/j.ecolind.2022.108990.

Seegers, B.N., Werdell, P.J., Vandermeulen, R.A., Salls, W., Stumpf, R.P., Schaeffer, B.A., Owens, T.J., Bailey, S.W., Scott, J.P., Loftin, K.A., 2021. Satellites for long-term monitoring of inland U.S. lakes: the MERIS time series and application for chlorophyll-a. Remote Sens. Environ. 266, 112685. https://doi.org/10.1016/j.rse.2021.112685.

Shahvaran, A., 2024. Chlorophyll-a Mapping in a Large Lake using Remote Sensing Imagery: a Case Study of Western Lake Ontario. Science of The Total Environment 968, 178881. https://doi.org/10.1016/j.scitotenv.2025.178881.

Shahvaran, A.R., Kheyrollah Pour, H., Binding, C., Van Cappellen, P., 2025. Mapping satellite-derived chlorophyll-a concentrations from 2013 to 2023 in Western Lake Ontario using Landsat 8 and 9 imagery. Sci. Total Environ. 968, 178881. https://doi.org/10.1016/j.scitotenv.2025.178881.

Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. Limnol. Oceanogr. 50, 237–245. https://doi.org/10.4319/lo.2005.50.1.0237.

Smith, J., Sutula, M., Bouma-Gregson, K., Van Dyke, M., 2021. California Water Boards' Framework and Strategy for Freshwater Harmful Algal Bloom Monitoring: Full Report with Appendices (Techincal No. 1141.B). California State Water Resources Control Board, Sacramento, CA.

Streib, L.C., Stone, J.R., Lyon, E.C., Quang, H.H., Yeager, K.M., Zimmerman, S.R.H., McGlue, M.M., 2021. Anthropogenic climate change has altered lake state in the Sierra Nevada (California, USA). Glob. Chang. Biol. 27, 6059–6070. https://doi.org/10.1111/gcb.15843.

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

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

Verhagen, J.H.G., 1994. Modeling phytoplankton patchiness under the influence of wind-driven currents inlakes. Limnol. Oceanogr. 39, 1551–1565. https://doi.org/10.4319/lo.1994.39.7.1551.

Vettorazzo, S., Boscaini, A., Cerasino, L., Salmaso, N., 2024. From small water bodies to lakes: exploring the diversity of freshwater bacteria in an Alpine Biosphere Reserve. Sci. Total Environ. 954, 176495. https://doi.org/10.1016/j.scitotenv.2024.176495.

Vincent, W.F., 2007. Cold Tolerance in Cyanobacteria and Life in the Cryosphere. In: Seckbach, J. (Ed.), Algae and Cyanobacteria in Extreme Environments. Springer, Netherlands, Dordrecht, pp. 287–301. https://doi.org/10.1007/978-1-4020-6112-7_15.

Wang, S., Qin, B., 2025. Application of optical remote sensing in harmful algal blooms in lakes: a review. Remote Sens. 17. https://doi.org/10.3390/rs17081381.

Whitman, P., Schaeffer, B., Salls, W., Coffer, M., Mishra, S., Seegers, B., Loftin, K., Stumpf, R., Werdell, P.J., 2022. A validation of satellite derived cyanobacteria detections with state reported events and recreation advisories across U.S. lakes. Harmful Algae 115, 102191. https://doi.org/10.1016/j.hal.2022.102191.

World Health Organization, 1999. Toxic Cyanobacteria in Water: a Guide to their Public Health Consequences. Monitoring and Management. E & FN Spon, New York. https://www.who.int/publications/m/item/toxic-cyanobacteria-in-water-a-guide-to-their-public-health-consequences-monitoring-and-management.

Wu, J., Cao, Y., Wu, S., Parajuli, S., Zhao, K., Lee, J., 2025. Current Capabilities and challenges of Remote Sensing in monitoring Freshwater Cyanobacterial Blooms: a Scoping Review. Remote Sens. 17. https://doi.org/10.3390/rs17050918.

Wu, T., Qin, B., Zhu, G., Luo, L., Ding, Y., Bian, G., 2013. Dynamics of cyanobacterial bloom formation during short-term hydrodynamic fluctuation in a large shallow, eutrophic, and wind-exposed Lake Taihu, China. Environ. Sci. Pollut. Res. 20, 8546–8556. https://doi.org/10.1007/s11356-013-1812-9.

Wynne, T.T., Meredith, A., Stumpf, R., Briggs, T., Litaker, W., 2021. Harmful Algal Bloom Forecasting Branch Ocean Color Satellite Imagery Processing Guidelines: 2021 Update, NOAA Technical Memorandum NOS NCCOS 296. NOAA National Centers for Coastal Ocean Science. https://repository.library.noaa.gov/view/noaa/30906.

Wynne, T.T., Stumpf, R.P., Pokrzywinski, K.L., Litaker, R.W., De Stasio, B.T., Hood, R.R., 2022. Cyanobacterial Bloom Phenology in Green Bay using MERIS Satellite Data and Comparisons with Western Lake Erie and Saginaw Bay. Water 14. https://doi.org/10.3390/w14172636.

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

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

Zakhia, F., Jungblut, A.-D., Taton, A., Vincent, W.F., Wilmotte, A., 2008. Cyanobacteria in Cold Ecosystems. In: Margesin, R., Schinner, F., Marx, J.-.-C., Gerday, C. (Eds.), Psychrophiles: from Biodiversity to Biotechnology. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 121–135. https://doi.org/10.1007/978-3-540-74335-4_8.

Zeng, C., Binding, C.E., 2021. Consistent Multi-mission measures of inland water algal bloom spatial extent using MERIS. MODIS and OLCI. Remote Sens. 13. https://doi.org/10.3390/rs13173349.

Zhang, Y., Duan, H., Xi, H., Huang, Z., Tsou, J., Jiang, T., Liang, X.S., 2018. Evaluation of the influence of aquatic plants and lake bottom on the remote-sensing reflectance of optically shallow waters. Atmos. Ocean 56, 277–288. https://doi.org/10.1080/07055900.2018.1454295.