

University of Vermont

UVM ScholarWorks

UVM Honors College Senior Theses

Undergraduate Theses

2022

Analyzing Interannual Cyanobacteria Bloom Variability in Lake Champlain using Satellite Imagery

Mia Pienkowski
University of Vermont

Follow this and additional works at: <https://scholarworks.uvm.edu/hcoltheses>

Recommended Citation

Pienkowski, Mia, "Analyzing Interannual Cyanobacteria Bloom Variability in Lake Champlain using Satellite Imagery" (2022). *UVM Honors College Senior Theses*. 495.
<https://scholarworks.uvm.edu/hcoltheses/495>

This Honors College Thesis is brought to you for free and open access by the Undergraduate Theses at UVM ScholarWorks. It has been accepted for inclusion in UVM Honors College Senior Theses by an authorized administrator of UVM ScholarWorks. For more information, please contact scholarworks@uvm.edu.

Analyzing Interannual Cyanobacteria Bloom Variability in Lake Champlain using Satellite Imagery

Mia Pienkowski

University of Vermont

College of Engineering and Mathematical Sciences

Department of Computer Science and Department of Mathematics and Statistics

Defense Date: April 12, 2022

Advisor: Scott Hamshaw, Ph.D

Committee Member: Donna Rizzo, Ph.D

Committee Member: Andrew Schroth, Ph.D

Table of Contents

Introduction	4
Methods	10
Data	10
Results	22
Discussion	44
Conclusion	49
References	51
Appendix A	61
Appendix B	68

Introduction

Problem Statement

This research sought to help address the problem of the lack of certainty of the spatial patterns of blooms in Lake Champlain. This uncertainty comes from the dependence on in-situ data that only represents specific points within the lake. The overall goal of this project was to characterize spatiotemporal patterns of cyanobacterial blooms in Lake Champlain using data derived from satellite imagery. Using satellite imagery allowed for the analysis of cyanobacterial bloom patterns over the entire spatial range of blooms as opposed to the more limited range of field collected data. The datasets formed from the processed imagery could be helpful for future research studying algal blooms and associated water quality in Lake Champlain.

Background

Harmful Algae Blooms

Blue-green algae or cyanobacteria are a type of photosynthetic aquatic bacteria that can be found in water bodies globally (*Algae, Phytoplankton and Chlorophyll*, n.d.; Stanier & Cohen-Bazire, 1977). Bloom events can be characterized and classified by numerous variables including nutrient concentration, size, growth rate, and abundance. Harmful algal blooms (HABs) are caused by numerous species of cyanobacteria and defined by the potential of an algal bloom to become harmful to the ecosystem or human health (*Nutrients and Eutrophication*, 2019; Stanier & Cohen-Bazire, 1977). Cyanobacteria are the only known freshwater algae which releases toxins capable of harming human health, and also have been linked to other economic problems for local communities in the Lake Champlain Basin such as decreased

property values and suppressed tourism (*Nutrients and Eutrophication*, 2019; Sellner et al., 2003).

The trophic state index (TSI) is a common method of classifying an aquatic ecosystem's biotic and nutrient conditions by quantitatively partitioning water bodies into the categories oligotrophic, mesotrophic, and eutrophic, based on biological productivity and nutrient levels (Carlson, 1977); A typical lake will gradually go through all three phases as the lake fills in with sediment over time and sediment organic matter concentration increases over time in sediment. A eutrophic lake is the most susceptible to HABs because they are in the highest stage of biological productivity. Eutrophication is a natural process caused by an increasing abundance of nutrients in water over time as the lake naturally changes, but it can be accelerated by anthropogenic activities which cause nutrients to collect in water at an increased rate (*Nutrients and Eutrophication*, 2019).

In recent decades, international scale research has led to growing concern regarding HABs in coastal and inland environments globally: there is thought to be an unprecedented increase in frequency and severity of HABs due to a combination of and positive feedbacks between land-use and climate change, nutrient loading and HAB occurrence, duration and severity (Chorus & Bartram, 1999; Anderson et al., 2002; H. W. Paerl & Paul, 2012; Taranu et al., 2015; Huisman et al., 2018). Furthermore, the current rate of increase in anthropogenic eutrophication and HAB events is expected to worsen as more intense effects of climate change occur, including here in the Lake Champlain Basin (H. Paerl & Huisman, 2009; Huang et al., 2020; Zia et al. 2016).

The most common drivers of the development of HABs that have cyanobacteria present (cyanoHAB) are generally considered to be a collection of hydrological and meteorological variables including temperature, precipitation, wind speed, atmospheric pressure, solar radiation, and nutrient loadings - which have been shown to interact with other hydrological variables (Jia et al., 2019; H. Paerl & Huisman, 2009). Given the effect meteorological variables and weather events have on blooms, it is apparent that climate change affects bloom patterns as cyanobacterial growth increases with higher temperatures, higher atmospheric carbon dioxide, and more intense precipitation events (Melendez-Pastor et al., 2019). Nutrient input is generally considered to be one of the most important long term drivers of blooms, as research has shown a strong relationship between lake nutrient concentrations and high biomass algal blooms (Anderson et al., 2002).

Cyanobacteria have always been found in Lake Champlain (Watzin et al., 2006), but the presence of toxins produced by cyanobacteria were not discovered until 1999 (Rosen et al., 2001). Similar to lakes and other water bodies globally, Lake Champlain is experiencing an increase in cyanobacterial blooms and toxins (Isles et al., 2015). The bloom period in Lake Champlain typically ranges from late June to early October: when temperatures are higher and there are more hours of sunlight. The frequency and severity of blooms in Lake Champlain's more shallow bays, including Missisquoi Bay and St. Alban's Bay has been increasing in recent decades (Isles et al., 2015; Levine et al. 2012). The shallow bays are more affected than deeper bays like Malletts Bay because deeper bays have more water volume which allows the nutrients that feed algae to dilute more (*Lake Look*, 2020). Due to the variety of conditions across the segments of Lake Champlain, the trophic state levels of the lake are assessed in 12 segments, rather than for the entire lake (Figure 1). Data from 2012 shows that 4 of these segments, St.

Alban's Bay, Missisquoi Bay, South Lake A, and South Lake B, are eutrophic while the other 8 segments are mesotrophic (Xu et al. 2015a, b). These eutrophic states have become more frequent and severe in the past two decades (Smeltzer et al., 2009). Phosphorus (P) loading from runoff is believed to have had a significant impact on these intensifying conditions and it is thought to be the main nutrient associated with the recent cyanobacterial growth (Trescott & Park, 2013; Torbick & Corbiere, 2015).

Amongst other species of algae, cyanobacteria are capable of producing scum, odors, and toxins (Sellner et al., 2003) which threaten the aquatic ecosystem, public health, domestic and wildlife, and the economic security of local communities (Wynne & Stumpf, 2015). As water quality continues to decline and the anthropogenic drivers of cyanoHABS intensify, the environmental and socioeconomic issues caused by cyanoHABS become increasingly concerning (*Toward a Satellite-Based Monitoring System for Water Quality*, 2017.). Within aquatic ecosystems, cyanobacteria growth can cause increased turbidity which would lessen the growth of aquatic plants, cause changes in the food web, and cause depletion of oxygen (Kim et al., 2017; H. Paerl & Huisman, 2009). When cyanobacteria occur in large quantities and release toxins, they can cause a variety of human health issues including respiratory and skin irritation (Stewart et al., 2006), gastrointestinal distress, dermatitis, liver failure (Mishra et al., 2019), allergic response, and promotion of tumors (Chorus & Bartram, 1999). Additionally, those relying on drinking water from sources affected by cyanoHABS may have increased difficulty accessing water as challenges in treating the drinking water occur. (Carmichael & Boyer, 2016; Watson et al., 2016; Zamyadi et al., 2012). The toxins released during cyanoHAB events can also cause death and sickness in domestic and wild animals as well as livestock if the animals are exposed to the water (Carmichael, 2001; Carmichael & Boyer, 2016; Bowling et al., 2015;

Carmichael et al., 2001). Due to the numerous potential negative effects of cyanoHABs, many countries have water usage and management restrictions regarding cyanoHAB presence (Bowling et al., 2015).

As the frequency of toxic cyanobacterial blooms has risen in recent decades, the need for monitoring and risk assessment systems has increased (Watzin et al., 2006). Various classification systems have been developed in an attempt to regulate and minimize the harmful effects of cyanoHABs on human health. Based on these systems for identifying risk level, the relative effect of cyanoHABs on human health can be determined. In some of these systems that quantify the risk for humans, the health and wellbeing of non-human land animals is also considered. These systems are subjective and there is some variation in how risk is classified in various localities and scales. The risk to the aquatic ecosystem is generally not encompassed in these systems that determine land animal health risk, but in the trophic state classification of a lake.

Remote Sensing and Spatial Analysis

Chlorophyll-a is a photosynthetic pigment found in all algae (including cyanobacteria) which is used as a proxy for algal biomass (Kim et al., 2017; Carlson, 1977; Coskun et al., 2008). Phycocyanin is another photosynthetic pigment, though it is unique to cyanobacteria, allowing it to be used to differentiate cyanobacteria from other algae and plants containing chlorophyll-a in satellite imagery (Lunetta et al., 2015; Stumpf et al., 2016; Urquhart et al., 2017). Understanding algal biomass is important in assessing risk and bloom patterns because it often indicates the trophic state of a water body (Hunter et al., 2009). Traditional methods for measuring chlorophyll-a concentration are generally finding in-situ measurements from sensors placed at

various points throughout a water body and shipboard sampling (Trescott & Park, 2013). As technologies have advanced, remote sensing has been increasingly employed in the assessment of the trophic status of water bodies, measuring water quality parameters, and monitoring algal blooms (Dörnhöfer & Oppelt, 2016; Tyler et al., 2016; Cullen et al., 1997; Ruddick et al., 2001). The usage of remote sensing in algal bloom monitoring could provide information that cannot be abstracted from in-situ data; due to the nature of field sampling, field data does not have the potential to explain the spatial patterns of cyanobacterial blooms (Jia et al., 2019). The specific points where data is collected through field sampling are not necessarily representative of the entire water body and can indicate different conclusions than what remote sensing imagery of the larger water body segment may reveal (Isenstein & Park, 2014). This is especially true when considering water bodies with great spatial variability (Urquhart et al., 2017). Although the position of the points measured in traditional field sampling methods can be moved, some areas may be too difficult to access or too costly to access (Chen & Quan, 2012). In addition to the greater spatial resolution remote sensing datasets can provide, they are often less expensive and labor-intensive to collect (Kim et al., 2017; Jia et al., 2019). The color of cyanobacterial blooms, influenced strongly by the presence of chlorophyll-a, means that most of these blooms are observable from imagery obtained via multispectral sensors on satellites. (Cullen et al., 1997; Gómez et al., 2021).

The spatiotemporal distributions of bloom patterns are generally complex and not necessarily captured well with traditional monitoring (Melendez-Pastor et al., 2019). In several cases, blooms have begun to be monitored with remote sensing. In Lake Winnipeg, the spatiotemporal variability of blooms was analyzed using metrics derived from OLCI and MERIS (Medium Resolution Imaging Spectromete) imagery (Binding, 2018). In Taihu Lake, nitrogen

and phosphorus concentrations were estimated using Landsat Thematic Mapper imagery (Chen & Quan, 2012). In recreational and drinking water sources in Ohio and Florida, methods have been developed to assess bloom frequency, magnitude, and spatial extent using OLCI and MERIS imagery (Clark et al., 2017; Mishra et al., 2019; Urquhart et al., 2017). In Lake Erie, cyanobacteria index derived from hyperspectral imagery has contributed to a weekly report on water quality since 2015 (Vander Woude et al., 2019). There is a need for further exploration of techniques employing remote sensing imagery (Chang et al., 2012). This research has sought to utilize available output from the OLCI sensor on Sentinel-3 to assess the utility for studying cyanobacteria bloom occurrence, spatial distribution, and severity, over time, to supplement ongoing monitoring efforts derived from sensors and water sampling programs. It has also sought to develop a reproducible method to process the cyanobacteria satellite product derived from this OLCI imagery.

Methods

Data

Satellite Imagery

The primary dataset consisted of imagery of the Lake Champlain Basin obtained from 2016-present from the Ocean and Land Color Instrument (OLCI) on the Sentinel-3 satellite . OLCI is a spectrometer with 21 bands measuring solar radiation from Earth at a spatial resolution of 300m (*Copernicus Sentinel-3 Ocean and Land Colour Instrument (OLCI) - LAADS DAAC*, 2018). The data from 04/22/2016 - 06/30/2018 was already in the form of 7-day maximum composites which were produced by NOAA from individual images that were taken almost daily, as the original near-daily OLCI images were not available the HAB Data Explorer

at the time of this study (National Centers for Coastal Ocean Science [NCCOS], 2022). The data from after 06/30/2018 was in the form of individual photos that are taken almost daily.

A measurement called the Cyanobacteria Index (CI) developed by NOAA had already been calculated for some images in the dataset. CI is a measure of cyanobacteria (Vander Woude et al., 2019). Within the dataset, there are 4 images given for every image taken, with the 4 images showing different bands on OLCI: one represents the CI for all algae, one represents the CI for specifically cyanobacteria, one represents the CI for all non-cyanobacteria, and one which is true color. For the purposes of this research, only the images that represented the CI for specifically cyanobacteria were used.

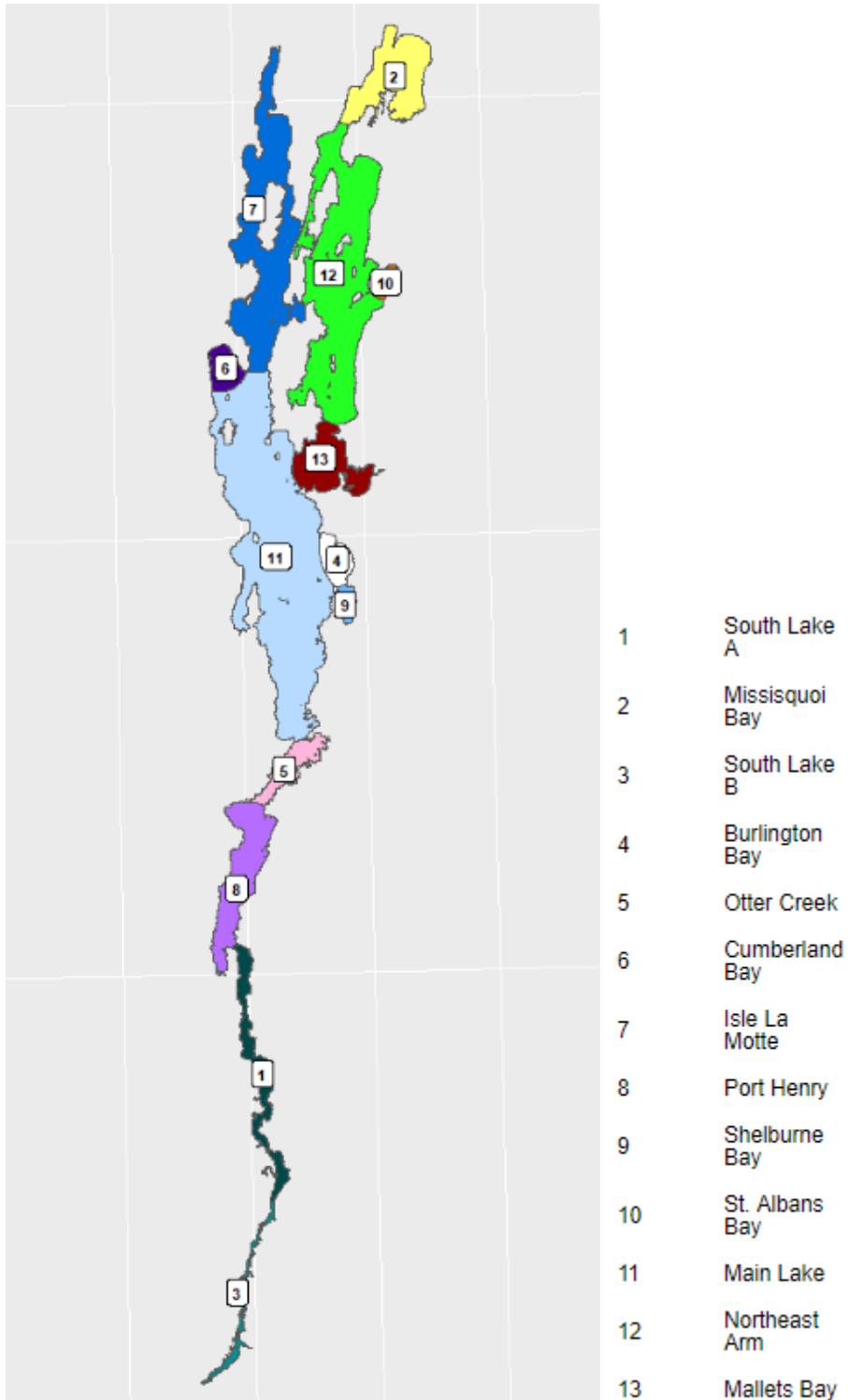
Images were stored on the National Centers for Coastal Ocean Science (NCCOS) HAB Data Explorer for approximately a year after the date they were taken on. At the time of downloading, HAB Data Explorer storage went back to 09/03/2020. In total, 867 images were bulk downloaded from the HAB Data Explorer using an R script (Bitterman, 2021). Images from 07/01/2018 - 09/02/2020 and 7-day maximum composites from 04/22/2016 - 06/31/2018 were obtained from Patrick Bitterman.

Lake Outline

A shapefile containing the border of lake champlain as well as the borders of 13 lake segments comprising Lake Champlain was downloaded from the Lake Champlain Basin Program Atlas (<https://atlas.lcbp.org/nature-environment/lake-champlain/>). These 13 segments, which are depicted in Figure 1, are South Lake A, Missisquoi Bay, South Lake B, Burlington Bay, Otter Creek, Cumberland Bay, Isle La Motte, Port Henry, Shelburne Bay, St. Albans Bay, Main Lake, Northeast Arm, and Mallets Bay.

Figure 1

Map of lake segments in Lake Champlain



Compositing Images

The images from July 2018 through December 2021 were composited using the tool “Create time series composite” from the ArcGIS Python toolbox, RS Tools in ArcMap (RS Tools User Guide). For each set of 7 consecutive images such that no image is in more than one set, this tool outputs one image, for which each pixel value represents the maximum of the several cyanobacteria index values from each of the images. A set of consecutive images that are composited can be made of images from different months. However, a composite cannot be made of images from different years.

Data Processing

All analysis was done using version 4.0.3 of the statistical software R. The composited images from April 2016 through December 2021 were stored in a raster stack using the raster package (<https://cran.r-project.org/web/packages/raster/index.html>).

The data was reclassified so that the value assigned to invalid pixel values was labeled as a missing value, NA (not available). This was done because the value for invalid pixels was on the same scale as the cyanobacteria index values for valid pixels and would have interfered with analysis of these valid CI values.

In the raw format, the pixel values are represented as digital numbers (DNs) on a scale of 1-249. The transformation given in Figure 2 from the readme file on the HAB Data Explorer was used to convert the DN to cyanobacteria index values, using the calc function in the raster package (<https://cran.r-project.org/web/packages/raster/index.html>).

Figure 2

Transformation from satellite digital number to cyanobacteria index value

$$10^{\left(\frac{3}{250*DN-4.2}\right)}$$

Note: DN stands for Digital Number

For certain analyses where individual lake segments were being considered, 13 unique raster stacks were derived from the original raster stack. Each of these 13 stacks were cropped to one of the 13 lake segments that were looked at in this research. This derivation was done using the crop and mask functions in the raster package

(<https://cran.r-project.org/web/packages/raster/index.html>). The shapefile of the lake segments was processed using the readOGR and spTransform functions in the rgdal package

(<https://cran.r-project.org/web/packages/rgdal/index.html>).

For analysis regarding cloud cover, the individual images were needed as they contained more information than the composited images. In their original form, the pixel values 0 and 250-255 uniquely indicated different types of invalid data. A pixel value of 253 indicated cloud cover which was used to analyze cloud cover in the dataset. Once the images were composited, all invalid pixels were labeled with 1 value as opposed to the 6 unique values, so the composited images could not be used for cloud cover analysis.

Raster Analysis

To better observe temporal and spatial trends of the cyanobacteria, raster sums and maxes were computed. The raster sum is a raster for which each pixel's value is the sum of all of the

values at that pixel from the rasters that were compiled. The raster max is a raster for which each pixel's value is the maximum of all of the values at that pixel from the rasters that were compiled. These calculations were performed using the `stackApply` function from the raster package (<https://cran.r-project.org/web/packages/raster/index.html>). The sums were chosen to be computed because they could highlight particular areas of the lake that either have one to a few very severe blooms or areas that were more consistently blooming across 2016-2021. The maxes were chosen to be computed because they could highlight areas that experienced comparatively severe bloom activity. Comparing the maxes and the sums also provides more information about what is seen in the sums. If the areas with high sums were different from the areas with high maxes, that could indicate that the areas where consistent blooms are appearing are different from area(s) that experienced the highest severity. The sum and max rasters were plotted using the `ggplot2` package, the `dplyr` package, and the `scico` package (<https://ggplot2.tidyverse.org/>, <https://dplyr.tidyverse.org/>, <https://github.com/thomas85/scico>). In the plots, the sums and maxes were plotted on top of the shapefile for the lake to highlight the proportion and area of the lake that is showing.

To further understand the trends shown in these sum and max rasters, the sums and maxes were computed over different time periods and segments. All of the different time periods used for these sum and max rasters were calculated for the whole lake, Missisquoi Bay, St. Albans Bay, and the Northeast Arm. Sum and max rasters were computed for each year such that only composites made of images from a particular year went into the sum and max calculations. They were also computed for a general bloom season time period (April - October) and not bloom season (November - March), for each year and for all years combined.

Cloud Cover Analysis

To analyze cloud cover, the number of pixels with the value 253 which was coded for cloud cover was counted for each image from 2020 for the entire lake. The number of pixels coded for cloud cover was then subtracted from the total number of pixels to give the number of pixels not covered by clouds. The number of pixels not covered by clouds was then divided by the total number of pixels to give the percentage of pixels not covered by clouds. This data was plotted using the package ggplot2 (<https://ggplot2.tidyverse.org/>).

Classifying Blooms

To assess the bloom patterns temporally, datasets of blooms were derived from the composites. Three datasets of blooms were derived, one specific to Missisquoi Bay, one specific to St. Albans Bay, and one specific to the Northeast Arm. These three segments were focused on based on data availability, as well as interest in bloom behavior in these systems due to their ongoing eutrophication (Isles et al., 2017). For the derivation of each of these datasets, the stacks of composites cropped to the different segments were processed. For each of the segments, an intermediary dataset was produced. These datasets' records were the 297 cropped composites. For the purposes of analysis, records were dated by the first image that was in the composite. Each record was labeled twice, based on 2 different sets of criteria, as 'appears to be a bloom', 'appears to not be a bloom', or 'inconclusive'. For both sets of criteria, the same cyanobacteria index threshold of .001 was used. This use of this threshold value was determined based on the relation in figure between cyanobacteria index and cell abundance in cells/mL and World Health Organization (WHO) risk assessment levels in cells/mL for harmful algae blooms. The threshold value .001 corresponds to the cell abundance value 99,999 cells/mL, the beginning

value of the range for high risk according to WHO (Lunetta et al., 2015). This relation has been validated on Lake Champlain. This process could be repeated with other threshold values.

The first set of criteria labeled a composite as ‘appears to be a bloom’ if the composite had any pixel that was at or above the threshold .001 for cyanobacteria index. This criteria had the potential to identify potential blooms that were only represented by a few spatially separated pixels due to cloud cover. However, it also classified composites that shouldn’t be considered as ‘appears to be a bloom’ as ‘appears to be a bloom’. The other set of criteria labeled a composite had a spatial threshold of 4 adjacent pixels. A composite was labeled as ‘appears to be a bloom’ if there were 4 adjacent pixels at or above the threshold (where adjacent meant the 8 pixels touching a given pixel. The threshold value of 4 could be changed.

There was the potential that a given composite could have multiple distinct groups of at least 4 adjacent pixels. This was especially true if a composite was not cropped and represented the entire lake. For purposes of visual analysis, this process was mostly performed on cropped composites. This made it so that the bloom patterns in small segments were able to be observed.

While the distinct groups of pixels could represent a singular bloom appearing this way due to invalid data or patchy bloom geometry, it could have also identified 2 separate bloom events occurring at the same time. The clump function from the raster package was used to identify distinct patches within a given composite so that the number of patches in each composite with a size of at least 4 pixels could be kept track of

(<https://cran.r-project.org/web/packages/raster/index.html>).

A feature indicating whether or not a given composite had at least one valid pixel was also developed for the intermediary dataset. This feature gave one method of assessment of how much usable data a given lake segment had. This feature also provided a way to differentiate

between composites that were labeled as not appearing to be a bloom, and composites whose label cannot be concluded due to lack of data.

The records in the intermediary dataset labeled as blooms according to the 4 pixel threshold were labeled as start dates, end dates, or mid bloom dates, by parsing the dataset and finding each composite placed in a sequence of composites labeled as appearing to have blooms. Composites for which the composites directly before and after them were labeled as not appearing to be blooms were labeled as start dates. The graphs in figures were made from these datasets using the packages `ggplot2`, `dplyr`, and `scico` (<https://ggplot2.tidyverse.org/>, <https://dplyr.tidyverse.org/>, <https://github.com/thomasp85/scico>).

Bloom Metrics

Bloom Start and End Dates

Three datasets of blooms, one for Missisquoi Bay, one for St. Albans Bay, and one for the Northeast Arm, were derived from parsing the intermediary datasets which contained dates and bloom categories for each of the composites. A bloom was considered to have been initiated at a given composite if the previous two composites were labeled as inconclusive or appeared to not be a bloom, and if either of the next two composites were labeled as appearing to be a bloom. This means that some composites that were labeled as appearing to be a bloom were not included in the bloom dataset. This rule was used to reduce false positives and variability in metrics that would be measured from the blooms. Once a bloom was considered initiated at a given composite, each of the following composites were considered part of that bloom if they were labeled as appearing to be a bloom, or if either of the two composites following were labeled as appearing to have a bloom. This allowed for a 2 composite gap in between 2 composites labeled

as having a bloom without starting a new bloom instance. This was done to account for missing data or weather events. A bloom was then considered to have senesced if the 3 composites following a given composite were labeled as inconclusive or as not appearing to be a bloom. For the composites 1, 2, 295, 296, and 297, special cases were used because there are not 2 previous composites or 3 following composites.

Bloom Duration

To keep track of and quantify the duration of blooms, a temporary dataset held the records of each of the composites that went into a given bloom. Once the bloom was considered to have senesced, the number of composites in the temporary dataset that were labeled as appearing to have a bloom could be counted. The duration could then be derived from this by multiplying by 7, since the composites are composed of 7 images.

Bloom Maximum Area

The maximum area for a bloom was calculated by finding the number of pixels included in a bloom patch in each composite that is part of a bloom, and keeping track of the highest value. To make the spatial extent of a bloom more comparable across segments, the maximum area was divided by the total valid area of the given lake segment. Total valid area of a lake segment was calculated by subtracting the number of pixels that did not show a valid data value in any of the composites from the total number of pixels in that lake segment.

Bloom Severity

The maximum severity for a bloom was calculated by finding the highest cyanobacteria index value for each composite that is part of a bloom, and keeping track of the highest value.

The tables showing these metrics were made using the flextable package and dplyr (<https://davidgohel.github.io/flextable/reference/index.html> , <https://dplyr.tidyverse.org/>).

Bloom Comparison Across Segments

To assess how many dates of blooms the lake segments had in common and not in common, portions of the composites labeled as having blooms were compared. Considering the three lake segments Missisquoi Bay, St. Albans Bay, and the Northeast Arm, the composites labeled as having blooms were split up into 6 categories: there was only a bloom in Missisquoi Bay, there was only a bloom in St. Albans Bay, there was only a bloom in the Northeast Arm, there was a bloom in both Missisquoi Bay and St. Albans Bay and not in the Northeast Arm, there was a bloom in both Missisquoi Bay and the Northeast Arm and not in St. Albans Bay, there was a bloom in both St. Albans Bay and the Northeast Arm and not in Missisquoi Bay, and there was a bloom in all three of the segments.

Assessment of Quantity of Non-Usable Data

It was important to quantitatively assess how much of a lack of data there was in each of the segments to understand the limitations of the dataset and the data processing implemented. In order to do this, the percent of pixels that never had a valid data value was calculated for each segment. Additionally, the percent of composites from April through November (generally bloom season) that did not have any valid data values was calculated for each segment.

The number of pixels that never had a valid data value for each segment was calculated by keeping a tracker matrix of the same shape as the matrices representing the rasters for each segment. This tracker matrix began having only missing values. Every composite per segment was then looped through, and for each composite, for every pixel that had a valid value, the

corresponding pixel of the tracker matrix was changed from having a value of NA to 1. After this process was completed, the number of pixels still having a value of NA in the tracker matrix was the number of pixels who never had a valid data value.

The number of composites per segment from April to November that did not have any valid data values was computed by first filtering the composites to only include those from April to November using the dplyr package (<https://dplyr.tidyverse.org/>). Then, each composite per segment was looped through. For each composite, every pixel's value was put into a vector. The vector was then parsed until a value that was valid appeared. If a value that was valid did not appear and the vector was completely parsed, then that composite was marked as not having any valid data.

Results

Assessment of Quantity of Non-Usable Data

Assessment of the lack of data in the dataset as well as analysis led to focusing on some segments more than others. Eutrophic segments of the lake like Missisquoi Bay, St. Albans Bay, and Shelburne Bay were of interest to compare amongst each other and to non-eutrophic segments. Missisquoi Bay had the lowest percentage of invalid composites during bloom season, with 22.01% (Table 1). The Northeast Arm, a non-eutrophic segment where concerns of gradual climate change induced eutrophication exist (Isles et al. 2017), had the second lowest percentage, at 35.41% (Table 1). St. Albans Bay had the 5th largest percentage, at 68.42% and Shelburne Bay had the largest percentage at 88.52%. Shelburne Bay also had the largest percentage of pixels that never had a valid value, at 54.72%. In addition to Shelburne Bay, Burlington Bay, Otter Creek, and Port Henry all were on the larger end for both the percentages

of invalid composites and the percentages of invalid pixels. St. Albans Bay had 0% invalid pixels, Missisquoi Bay had 0.56%, and the Northeast Arm had 32.45%.

Table 1

Percent of composites from April - November that had no valid data, grouped by lake segment

Segment	Percent of Composites
Missisquoi Bay	22.01%
Northeast Arm	35.41%
Mallets Bay	40.67%
Isle La Motte	43.54%
Main Lake	50.24%
South Lake A	54.01%
South Lake B	64.11%
Port Henry	68.42%
St. Albans Bay	68.42%
Otter Creek	72.73%
Cumberland Bay	80.38%
Burlington Bay	82.78%
Shelburne Bay	88.52%

Table 2

Percent of pixels that never had a valid value, grouped by lake segment

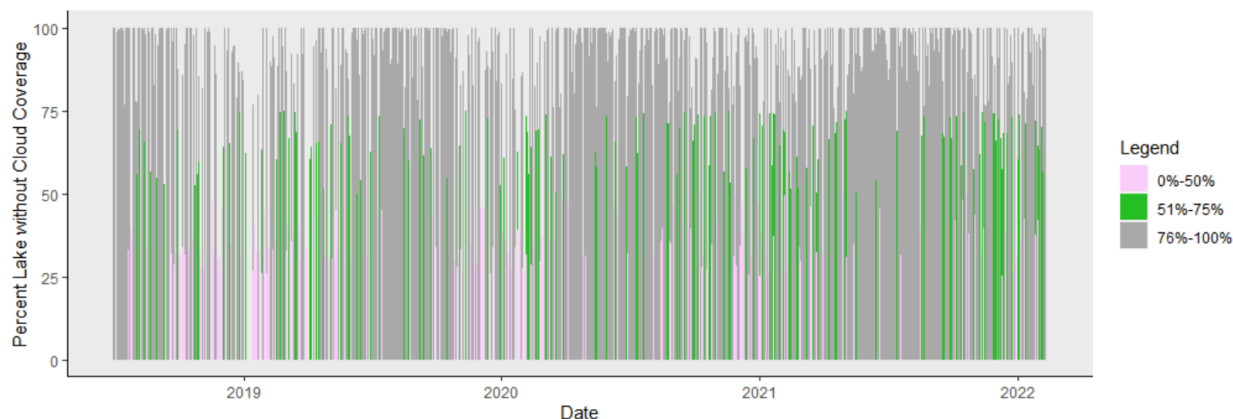
Segment	Percent of Pixels
St. Albans Bay	0.00%
Missisquoi Bay	0.05%
South Lake A	5.13%
Cumberland Bay	14.80%
South Lake B	15.29%
Isle La Motte	28.82%
Main Lake	31.06%
Northeast Arm	32.45%
Otter Creek	33.12%
Mallets Bay	35.15%
Port Henry	36.12%
Burlington Bay	37.00%
Shelburne Bay	54.72%

Cloud Cover Analysis

Seen in Figure 3, many of the images that went into the composites had at least 75% of pixels that were not classified as cloud cover. There are still a lot of images, though less, that had 25%-50% of pixels classified as cloud cover.

Figure 3

Percent of whole lake not covered by clouds per day from 07/01/2018 - 12/31/2021



Note. Each image from 07/01/2018 - 12/31/2021 is represented by a bar whose height is determined by the percent of pixels in the image that are not classified as cloud cover or land. The bars are colored by groups of percents.

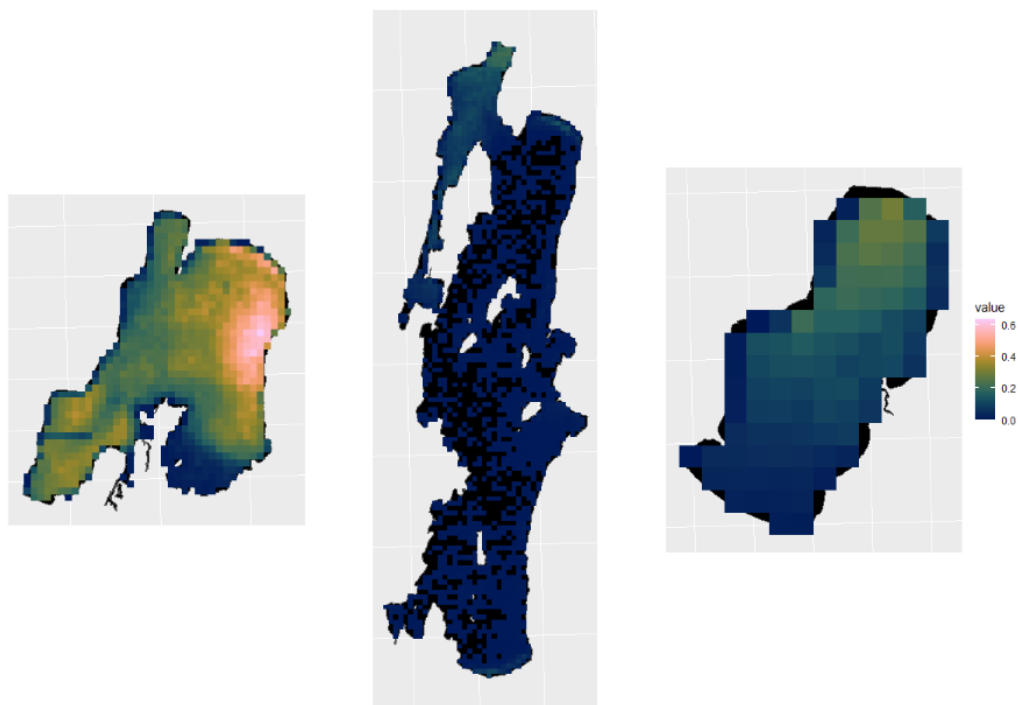
Raster Analysis and Comparison

The overall goal of the analysis was to be able to assess and gain insight regarding how the trends in cyanobacteria blooms compared spatially and temporally across the segments comprising Lake Champlain.

Raster summation and max plots were calculated to investigate general trends in the strongest bloom areas both within the three segments across years, and across the lake segments. As shown in Figure 4, Missisquoi Bay had the highest summed cyanobacteria index values. Missisquoi Bay also had the highest non-summed cyanobacteria index values (Figure 5). The smallest cyanobacteria index value in the maximum calculations was in the Northeast Arm (Figure 5).

Figure 4

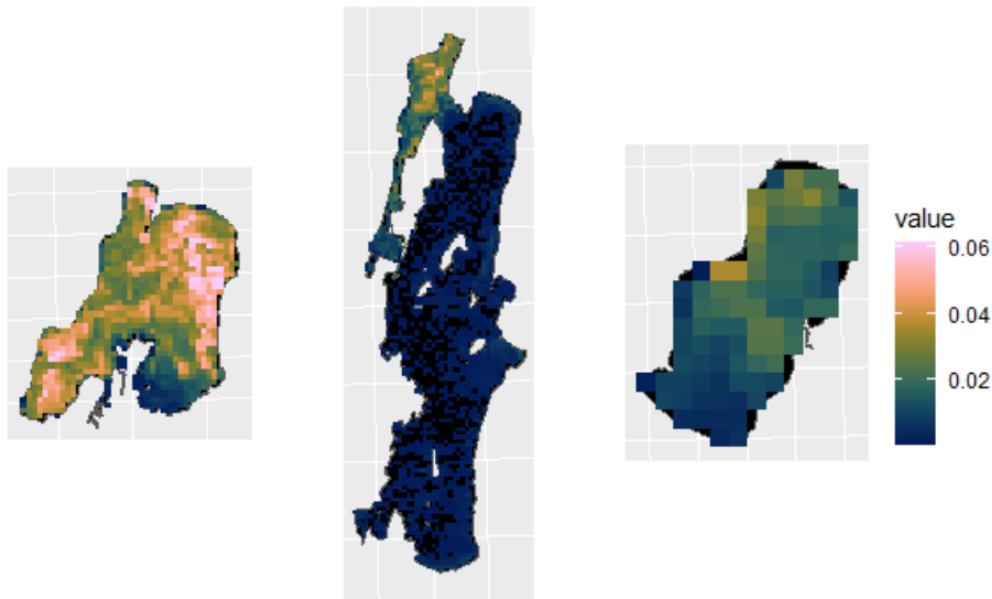
Cyanobacteria Index summations for composites from 04/22/2016 - 12/31/2021 for Missisquoi Bay, the Northeast Arm, and St. Albans Bay



Note. The segments appear in the order from left to right: Missisquoi Bay, the Northeast Arm, St. Albans Bay.

Figure 5

Cyanobacteria Index maximum calculations for composites from 04/22/2016 - 12/31/2021 for Missisquoi Bay, the Northeast Arm, and St. Albans Bay



Note. The segments appear in the order from left to right: Missisquoi Bay, the Northeast Arm, St. Albans Bay.

Sum and maximum calculations were computed for certain time ranges to temporally assess bloom trends within lake segments. St. Albans Bay (Figure 1A, Figure 2A), Missisquoi Bay (Figure 3A, Figure 4A), and the Northeast Arm (Figure 5A, Figure 6A) all had the highest individual and summed cyanobacteria index values in 2016. For St. Albans Bay, The lowest maximum summed value for any year was in 2020 (Figure 1A), for Missisquoi Bay, the lowest

maximum summed value was in 2019 (Figure 3A), and for the Northeast Arm, the lowest maximum summed value was in 2019 (Figure 5A).

In 2017, the Northeast Arm had the most bloom activity in a small section that connects to Missisquoi Bay (Figure 5A, 15). The majority of the bloom activity in Missisquoi Bay in 2017 happened in the northeastern region of the segment, and the least severe bloom activity occurred in the region closest to the Northeast Arm (Figure 3A).

In 2018, the second highest amount of variability in the Northeast Arm (after 2016) occurred in the northernmost region of the segment, where it extended south and west within that region (Figure 5A, Figure 6A). Missisquoi Bay had higher maximum cyanobacteria index values in its southernmost region compared to 2017 and 2019 (Figure 4A). Specifically, it had higher values than all other years aside from 2016 on the edges of the southernmost region, where it connects to the Northeast Arm (Figure 4A).

The highest summed (Figure 5A) and individual (Figure 6A) cyanobacteria index values occurred within all of the Northeast Arm south of its northernmost region in 2020. In Missisquoi Bay in 2020, the highest summed values were in the southernmost region of the segment. Additionally, in comparison to all other years, Missisquoi Bay had smaller summed cyanobacteria index values in its eastern region (Figure 3A).

Similarly to in 2020, in 2021, the Northeast Arm had a larger spread of higher summed (Figure 5A) and individual (Figure 6A) cyanobacteria index values across the segment as compared to 2016, 2017, 2018, and 2019. In 2021 in Missisquoi Bay, the southernmost region had similar values as in 2020, although it did have smaller summed (Figure 3A) and individual (Figure 4A) values on the western edge of this region. However, the eastern region had much higher summed (Figure 3A) and individual (Figure 3A) values than in 2020. In St. Albans Bay

in 2021, the highest summed (Figure 1A) and individual (Figure 2A) values spread further down from the northern region similarly to in 2016 and 2017, though with smaller values. There were similar spreads and low maximum values in St. Albans Bay in 2019 and 2020, although there were higher summed values in 2019 (Figure 2A, Figure 3A). In all years, St. Albans Bay experiences the highest values in its northern region. The eastern region of the Northeast Arm that connects to St. Albans Bay has the largest spread in 2016 and in 2020.

In the 2 month period of April through May, Missisquoi Bay (Figure 8A) had some low summed cyanobacteria index values, especially in 2017, 2018, 2020, and 2021, while the other two segments had mostly invalid data. In June through July in St. Albans Bay (Figure 7A), valid data values only extended from the northernmost pixel to the southernmost pixel in 2020 and 2021. During this time period in Missisquoi Bay, there were low summed cyanobacteria index values (Figure 8A) spread across the eastern border of the segment in all years aside from 2017, with 2019 and 2020 having more activity on the northern part of this border and 2016 having more activity on the southern part of this border. Almost all of the data for the Northeast Arm for June through July was invalid (Figure 9A). All of the segments experienced their highest summed cyanobacteria index values in August through September every year (Figure 7A, Figure 8A, Figure 9A). The period of August through September contributed the most to the overall pattern for every segment, though the other two month periods contributed more for Missisquoi Bay (Figure 8A) than for the Northeast Arm (Figure 9A) and St. Albans Bay (Figure 7A).

Bloom Metrics

St. Albans Bay

The average bloom start date of blooms for St. Albans Bay was the Julian day 217.83 (August 4th in leap years, August 5th in non-leap years). The average end date of blooms was the Julian day 262.5 (September 18th for leap years, September 19th for non-leap years).

St. Albans Bay experienced its longest bloom of approximately 70 days in 2021 from late July to early October, which is 1.79 standard deviations away from the average duration (Table 3). This bloom was also the earliest starting bloom and latest ending bloom, beginning on July 23rd, which is 1.18 standard deviations away from the average start date (Table 3). It ended on October 2nd, which is 1.29 standard deviations away from the average end date (Table 3).

The shortest bloom length in St. Albans Bay of approximately 35 days occurred twice (Table 3). The latest starting bloom was one of these two shortest blooms, beginning in late August and ending in late September in 2019 (Table 3). This bloom's start date was the most different from the average start date, at 1.81 standard deviations away. This 2019 bloom was also the bloom with the smallest maximum spatial extent, with 52.12% of pixels being covered, making it 1.79 standard deviations away from the average maximum spatial extent (Table 3).

The other shortest bloom occurred from early August to early September in 2016, and had the earliest end date, at September 2nd, which is 1.7 standard deviations from the average end date (Table 3). This bloom had the highest maximum spatial extent, with 84.04% of pixels being covered, making it 1.16 standard deviations from the average maximum spatial extent (Table 3).

Table 3

Blooms in St. Albans Bay from 04/22/2016 - 12/31/2021

Start Date	End Date	Maximum Percent of Total Area	Duration
2016-08-05	2016-09-02	84.04%	35
2017-08-06	2017-09-17	77.66%	42
2018-08-05	2018-09-23	74.47%	49
2019-08-27	2019-09-24	52.13%	35
2020-07-29	2020-09-17	69.15%	49
2021-07-23	2021-10-02	71.28%	70

Missisquoi Bay

In Missisquoi Bay, one bloom event was tracked outside of the normal bloom season. This event could have been falsely detected as a bloom from the imagery, as it is unlikely a bloom occurred in March, April, and May in Missisquoi Bay. In 2020, a bloom lasted from mid March to early May (Table 4). This bloom was the shortest bloom event Missisquoi Bay experienced and was 1.22 standard deviations away from the average duration. It was also the bloom reaching the smallest maximum spatial extent and the most deviant from the mean maximum spatial extent, at 1.84 standard deviations from the average (Table 4). This bloom was also the smallest reaching bloom, with a maximum 6.6% of pixels covered (Table 4). It was 1.84 standard deviations less than the average maximum spatial extent. To more accurately compare start and end dates across the blooms occurring in the typical bloom season, this bloom was excluded from further analysis.

The average bloom start date of blooms occurring during the bloom season for Missisquoi Bay was the Julian day 207.33 (July 25th in leap years, July 26th in non-leap years). The average end date of blooms occurring during the blooms season was the Julian day 290.5 (October 16th for leap years, October 17th for non-leap years).

Missisquoi Bay experienced its longest bloom of approximately 119 days in 2021 from early July to late October (Table 4). This bloom also had the earliest start date of the blooms occurring during bloom season, occurring on July 2nd, approximately 25 days earlier than the average start date and 1.24 standard deviations away from the mean (Table 4). This bloom ended on October 23rd, at .16 standard deviations away from the average end date (Table 4). Thus the longest bloom had an atypical early start date and a typical end date.

The shortest bloom during the bloom season occurred in 2017 from late August to early October, lasting approximately 42 days (Table 4). This bloom was also the latest starting bloom, and began on August 27th, approximately 31 days after the average start date and 1.61 standard deviations away from the mean (Table 4). It ended on October 8th, which was .28 standard deviations away from the average end date (Table 4). Thus, it had an atypically late start date and a typical end date.

The second longest bloom of approximately 112 days occurred in 2020 from late July to mid December (Table 4). This bloom occurred on July 29th, approximately 4 days after the average start date (Table 4). It ended on December 10th, approximately 55 days after the average end date, and 1.61 standard deviations away (Table 4). So, this second longest bloom had a typical start date and an atypical end date.

The bloom reaching the largest maximum spatial extent occurred from late July to mid October in 2016, covering 87.43% of pixels, making it .79 standard deviations higher than the

average maximum spatial extent (Table 4). Its duration was typical, at approximately 84 days and .08 standard deviations away from the average duration (Table 4).

The bloom reaching the smallest maximum spatial extent occurred from early July to late August 2019, covering 33.78% of pixels, making it 1.95 standard deviations below the average maximum spatial extent (Table 4). Its duration was the second lowest; at approximately 56 days, it was .85 standard deviations away from the average duration (Table 4).

Table 4

Blooms in Missisquoi Bay from 04/22/2016 - 12/31/2021

Start Date	End Date	Maximum Percent of Total Area	Duration
2016-07-29	2016-10-14	87.43%	84
2017-08-27	2017-10-08	71.16%	42
2018-07-29	2018-10-21	84.51%	77
2019-07-09	2019-08-27	33.78%	56
2020-03-18	2020-05-06	6.62%	35
2020-07-29	2020-12-10	75.53%	112
2021-07-02	2021-10-23	79.57%	119

The Northeast Arm

In the Northeast Arm, two bloom events were tracked outside of the normal bloom season. In 2020, a bloom lasted from late March to mid April (Table 5). In 2021, a bloom lasted from mid March to late April (Table 5). These two blooms were the shortest blooms that occurred in the Northeast Arm, with the one that occurred in 2020 being approximately 21 days,

7 days shorter than the 2021 bloom, making it 1.2 standard deviations below the average duration (Table 5). The 2021 bloom was .94 standard deviations from the average duration. The 2020 bloom was also the bloom with the smallest maximum spatial extent, with 2.08% of pixels being covered, making it .88 standard deviations away from the average maximum spatial extent (Table 5). The off-season bloom in 2021 had a relatively typical maximum spatial extent, covering 5.9% of pixels, it was .46 standard deviations away from the average (Table 5). It also had a higher maximum spatial extent than the bloom that occurred during the bloom season in 2021, which covered 3.43% of pixels (Table 5). To more accurately compare start and end dates across the blooms occurring in the typical bloom season, these blooms were excluded from further analysis.

The average bloom start date of blooms occurring during the bloom season for the Northeast Arm was the Julian day 230.67 (August 17th in leap years, August 18th in non-leap years). The average end date of blooms occurring during the bloom season was the Julian day 292.83 (October 18th for leap years, October 19th for non-leap years).

The Northeast Arm experienced its longest bloom of approximately 91 days from mid August to mid November in 2016 (Table 5). This bloom had the second largest maximum spatial extent, covering 16.12% of pixels (Table 5). This bloom had the latest end date, ending approximately 31 days after the average end date, making it 1.13 standard deviations away from the average (Table 5). It started 1 day after the average start date. Thus, the longest bloom had a typical start date and a late end date.

The shortest bloom that occurred during the Northeast Arm's bloom season lasted for approximately 35 days from early August to early September in 2019, making it 1.16 standard deviations away from the average duration (Table 5). It started on August 6th, .93 standard

deviations away from the average start date, and ended on September 3rd, 1.76 standard deviations away from the average end date (Table 5). This was the earliest ending bloom of the blooms that occurred during bloom season (Table 5). Thus, the shortest bloom had a typical start date and an early end date. This bloom had the second smallest maximum spatial extent, covering 3.54% of pixels at .88 standard deviations away from the average maximum spatial extent (Table 5).

The bloom covering the largest maximum spatial extent of 29.61% of pixels occurred from early August to early November in 2020, making it 1.76 standard deviations away from the average maximum spatial extent (Table 5). It was also the second longest bloom, lasting approximately 84 days (Table 5).

The bloom covering the smallest maximum spatial extent of 3.43% of pixels occurred from late August to late October in 2021, making it .89 standard deviations away from the average maximum spatial extent (Table 5).

Table 5

Blooms in the Northeast Arm from 04/22/2016 - 12/31/2021

Start Date	End Date	Maximum Percent of Total Area	Duration
2016-08-19	2016-11-18	16.12%	91
2017-09-10	2017-10-15	8.15%	42
2018-08-12	2018-10-14	12.70%	49
2019-08-06	2019-09-03	3.54%	35
2020-03-25	2020-04-15	2.07%	21
2020-08-05	2020-11-05	29.61%	84
2021-03-12	2021-04-23	5.90%	28
2021-08-27	2021-10-30	3.43%	70

Comparison Across Segments

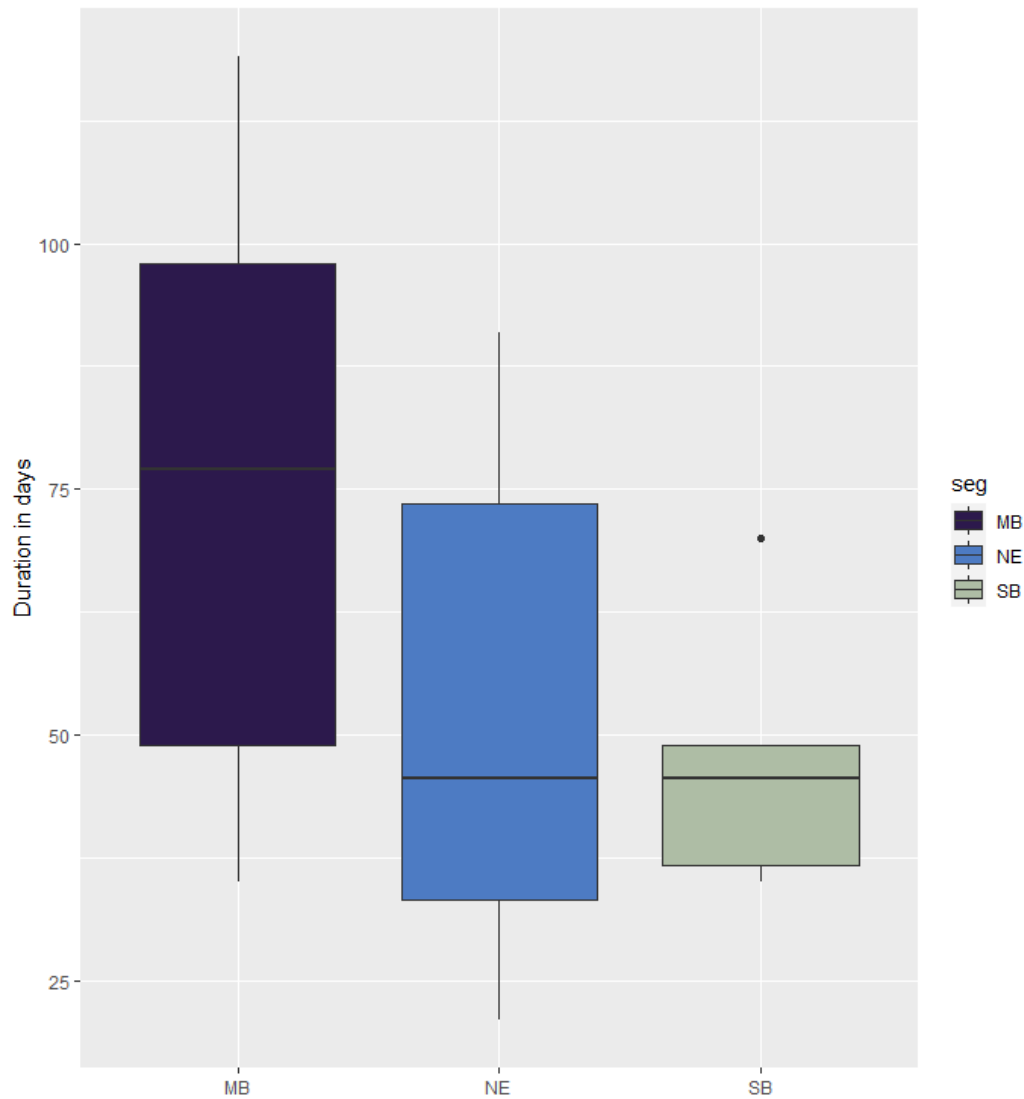
The highest durations were seen in Missisquoi Bay, while the lowest durations were seen in the Northeast Arm (Figure 6). Although the lowest durations were seen in the Northeast Arm, St. Albans Bay and the Northeast Arm had the same median duration of approximately 45.5 days (Figure 6). This is because a much broader range of durations was experienced in the Northeast Arm than in St. Albans Bay. So, although the lowest durations were in the Northeast Arm and not St. Albans Bay, the Northeast Arm also had blooms with larger durations than in St. Albans Bay (Figure 6).

The highest maximum spatial extent was seen in Missisquoi Bay and the lowest maximum spatial extent in the Northeast Arm (Figure 7). However, almost all of the maximum spatial extents from St. Albans Bay were close to the median value for Missisquoi Bay (Figure 7). Missisquoi Bay also experienced a lower minimum maximum spatial extent than St. Albans

Bay (Figure 7). The Northeast Arm had far smaller maximum spatial extents than the other two segments (Figure 7).

Figure 6

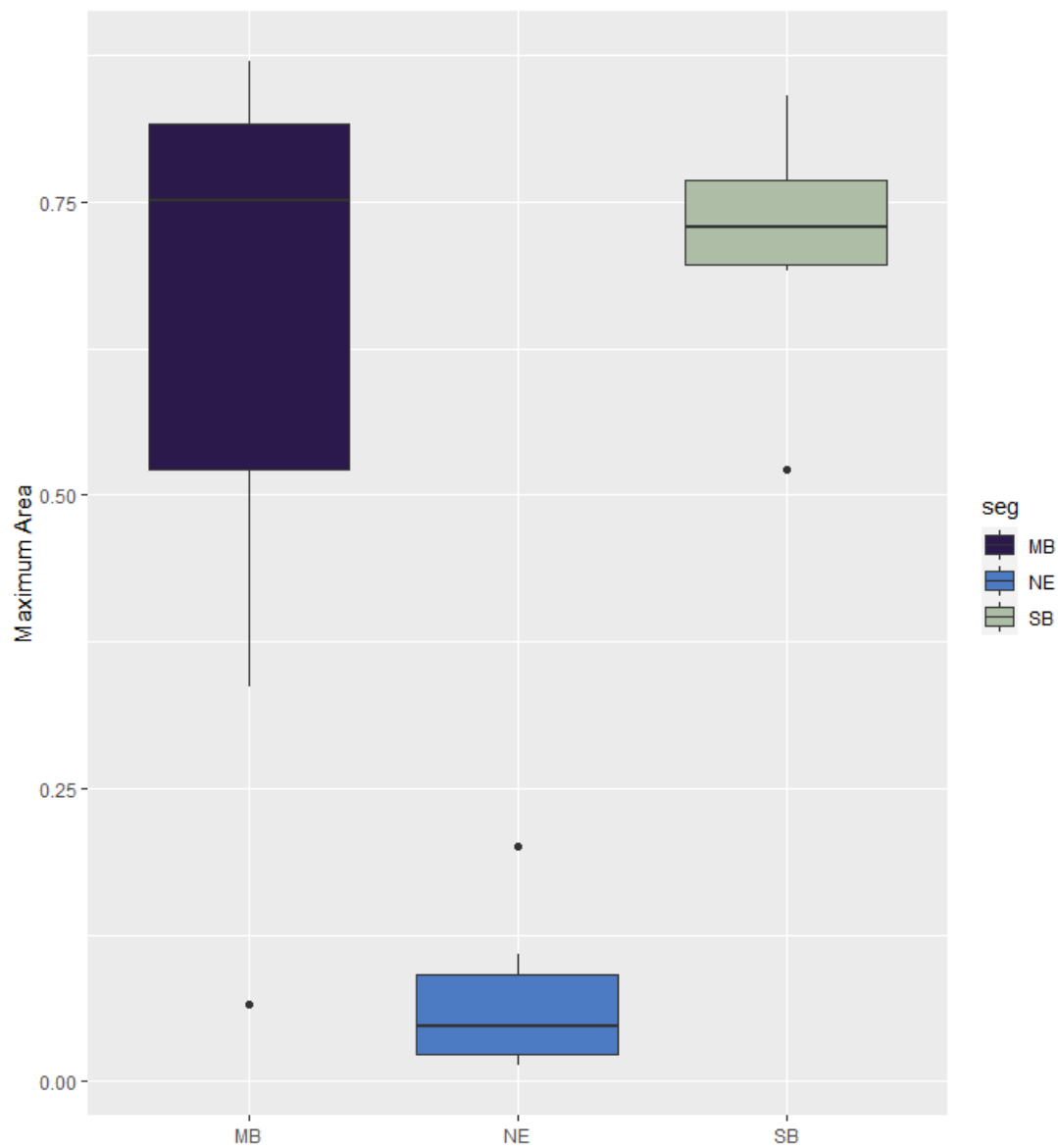
Box plot of duration in days grouped by segment



Note. MB stands for Missisquoi Bay, NE stands for Northeast Arm, and SB stands for St. Albans Bay.

Figure 7

Box plot of maximum spatial extent in percent of pixels per total area grouped by segment



Note. MB stands for Missisquoi Bay, NE stands for Northeast Arm, and SB stands for St. Albans Bay.

Summed and Maximum Cyanobacteria Index Values Analysis

In 2016, the highest values in the Northeast Arm occurred from early September to mid October (Figure 1B), with the highest total occurring in early September (Figure 2B). The highest values in Missisquoi Bay occurred from early August to mid October (Figure 3B), with the highest total occurring in early September (Figure 4B). The highest values in St. Albans Bay occurred from mid August to early September (Figure 5B), with the highest total occurring in mid-late August (Figure 6B).

In 2017, the highest values in the Northeast Arm occurred from mid September to late September (Figure 7B), with the highest total occurring in late September (Figure 8B). The highest values in Missisquoi Bay occurred from late August to early October (Figure 9B), with the highest total occurring in early October (Figure 10B). The highest values in St. Albans Bay occurred from mid to late September (Figure 11B), with the highest total occurring in mid September (Figure 12B).

In 2018, the highest values in the Northeast Arm occurred from late August to mid September (Figure 12B), with the highest total occurring in mid September (Figure 13B). The highest values in Missisquoi Bay occurred from mid August to mid September (Figure 15B), with the highest total occurring in early September (Figure 16B). The highest values in St. Albans Bay occurred from early August to late August (Figure 17B), with the highest total occurring in mid August (Figure 18B).

In 2019, the highest values in the Northeast Arm occurred in mid July and in mid August (Figure 19B), with the highest total occurring in mid July (Figure 20B). The highest values in Missisquoi Bay occurred from late July to mid August (Figure 21B), with the highest total occurring in mid August (Figure 22B). The highest values in St. Albans Bay occurred from early

September to late September (Figure 23B), with the highest total occurring in early September (Figure 24B).

In 2020, the highest values in the Northeast Arm occurred from mid August to mid September (Figure 25B), with the highest total occurring in late August (Figure 26B). The highest values in Missisquoi Bay occurred from early August to late September (Figure 27B), with the highest total occurring in mid-late September (Figure 28B). The highest values in St. Albans Bay occurred from early August to mid August (Figure 29B), with the highest total occurring in mid August (Figure 30B).

In 2021, the highest values in the Northeast Arm occurred from late August to mid September (Figure 31B), with the highest total occurring in late August (Figure 32B). The highest values in Missisquoi Bay occurred from early August to mid October (Figure 33B), with the highest total occurring in early September (Figure 54). The highest values in St. Albans Bay occurred from late July to early October (Figure 35B), with the highest total occurring in early August (Figure 36B).

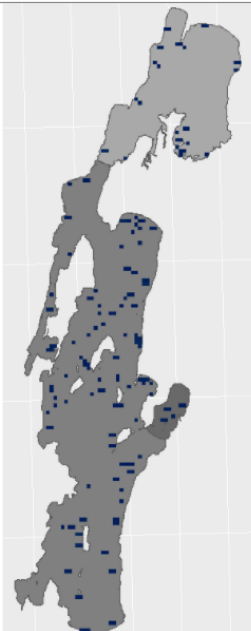
In the Northeast Arm, the time period in which the highest values occurred ranged from mid July to mid October, with some of these values occurring in September for every year except 2019. The highest total occurred in September in every year aside from 2019. In Missisquoi Bay, the time period in which the highest values occurred ranged from late July to mid October, with this time period beginning in August in every year but 2019. The highest totals occurred in September in every year except 2019 and 2017. In St. Albans Bay, the time period in which the highest values occurred ranged from late July to early October. The time period in which the highest totals occurred ranged from mid August to mid September.

Analysis of Locations of Blooms in Composites Labeled as Having Blooms

It is observed that of the composites labeled as blooms in only one segment, the most occurred in Missisquoi Bay, followed by the Northeast Arm (Table 6). Additionally the highest percentage of composites labeled blooms was the percentage of composites labeled as having blooms in Missisquoi Bay and the Northeast Arm and not in St. Albans Bay (Table 7). This is followed by the percentage of composites labeled as having blooms in all segments. The smallest overall was the percent of composites labeled as blooms that had blooms in the Northeast Arm and in St. Albans Bay and not in Missisquoi Bay.

Table 6

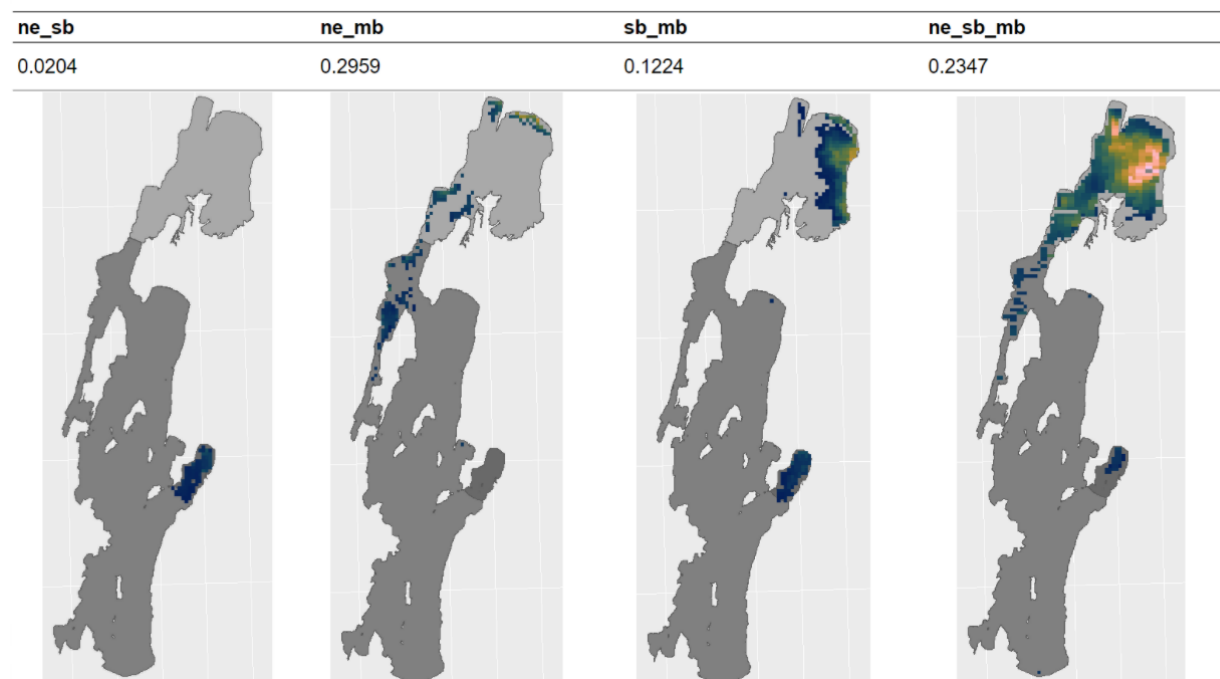
Proportion of composites labeled as blooms

only_ne	only_sb	only_mb
0.1327	0.051	0.1429
		

Note. Examples for `only_ne`, `only_sb`, and `only_mb` were taken from 12/04/2021, 09/10/2019, and 07/23/2019, respectively. The column `only_ne` represents the proportion of composites labeled as blooms in which a bloom was only found in the Northeast Arm. The column `only_sb` represents the proportion of composites labeled as blooms in which a bloom was only found in St Albans Bay. The column `only_mb` represents the proportion of composites labeled as blooms in which a bloom was only found in Missisquoi Bay.

Table 7

Proportion of composites labeled as blooms



Note. Examples for ne_sb, ne_mb, sb_mb, and ne_sb_mb were taken from 8/06/2017, 08/26/2018, 08/05/2018, and 09/02/2019, respectively. The column ne_sb represents the proportion of composites labeled as blooms in which a bloom was only found in the Northeast Arm and St. Albans Bay. The column ne_mb represents the proportion of composites labeled as blooms in which a bloom was only found in the Northeast Arm and Missisquoi Bay. The column sb_mb represents the proportion of composites labeled as blooms in which a bloom was only found in St. Albans Bay and Missisquoi Bay. The column ne_sb_mb represents the proportion of composites labeled as blooms in which a bloom was found in all three segments.

Discussion

Overall, the percentages of composites with no valid data were quite high, with the lowest at 22.01%, and the highest being 88.52%. Further investigation would be needed to determine why these numbers are as high as they are. It may be a result of the compositing process that was used. There does seem to be a pattern, as observed in the bloom graphs (Appendix A) that it is more common for a composite that is not during a bloom or bloom season to be considered inconclusive (having no valid data) than it is for composited from during a bloom or bloom season. This pattern would certainly bias the data such that segments that experience higher bloom frequency or magnitude would have more data than other segments. It is also concerning that there were such high percentages of pixels that never had a valid data value. With the compositing method that was used, it isn't possible to determine what invalid data classes these pixels had, but it would be worth further investigation to determine this. If they were mostly classed as adjacency and bordering land, then that would be of less concern. However, there was not a direct pattern that segments with larger perimeters had higher percentages of invalid pixels, as would be expected if this was the case. Figure 3 indicates that the reasoning there is such a lack of valid data is not only cloud cover. It is interesting that Shelburne Bay had such a lack of data when there is known to be bloom activity in Shelburne Bay, considering the pattern that occurred in other segments, that valid data occurred more frequently when there was bloom activity.

Based on results and data availability, the comparison between the two eutrophic bays Missisquoi Bay and St. Albans Bay, to the non-eutrophic Northeast Arm became focal.

As seen in Figure 4, Missisquoi Bay appeared to have the overall most severe blooms, especially compared to the Northeast Arm, which is consistent with other studies (e.g. Isles et al.

2017). This is highlighted by the lack of variability in color on the plot of the Northeast Arm, indicating that it has a smaller range of cyanobacteria index values, and that these values are generally much smaller than they are for Missisquoi Bay (Figure 4). The northernmost region of the Northeast Arm displayed the most variability within the segment, indicating that the most severe blooms in the Northeast Arm are coming from Missisquoi Bay (whose southernmost region connects to this region of the Northeast Arm) (Figure 4).

The maximum calculation (Figure 5) of Missisquoi Bay shows that there was at least one bloom event with some of the highest severity levels across the segments in the southwest region that would be the potential source of high bloom numbers in the adjacent northern end of the Northeast Arm. Comparing the maximum calculation (Figure 5) and the summation calculation (Figure 5) indicates that the northeast region of the segment experienced more frequent bloom events than other sections of the bay. These analyses of the satellite data illustrate the power of this spatially resolved dataset in identifying spatial patterns of bloom dynamics that would be evident from singular water sampling events.

Comparing bloom severity in St. Albans Bay and the Northeast Arm, they appear to have similar maximums of and range in severity levels, but a much larger portion of the Northeast Arm is on the lower end of the range than in St. Albans Bay (Figure 4, Figure 5). Additionally, the region of the Northeast Arm which is near St. Albans Bay does not appear to have a higher severity than the rest of the segment, indicating that St. Albans Bay influences bloom activity in the Northeast Arm on a smaller scale than Missisquoi Bay does (Figure 4).

The maximum calculation for the Northeast Arm showed that almost all of the variability, as well as the highest cyanobacteria index values were located in the northernmost region of the segment that connects to Missisquoi Bay (Figure 5). This, along with the lack of variability and

high values in the eastern region of the Northeast Arm that connects to St. Albans Bay (Figure 5) complements the findings from the summation calculations (Figure 4) that the Northeast Arm bloom activity is more influenced by Missisquoi Bay than St. Albans Bay. There was an instance in 2020 in which a lower severity bloom was distributed across much of the Northeast Arm (Figure 5A). This event could be representative of the Northeast Arm blooming independently of Missisquoi Bay or St. Albans Bay, or of the influence of these two bays on the Northeast Arm reaching a higher spatial extent.

It appears that St. Albans Bay had the least bloom activity in 2019 and 2020 (Figure 1A, Figure 2A). In both years, there is a similar spread and maximum of maximum individual values, but there are higher summed values in 2019. This indicates that there was more bloom activity in 2019 than in 2020, but not more severe bloom activity. In comparison, both the Northeast Arm (Figure 5A, Figure 6A) and Missisquoi Bay (Figure 3A, Figure 4A) appear to have had the least bloom activity, in 2019. Additionally, it appears that in all 3 segments, the highest number of pixels that never had a valid value occurred in 2019. This suggests that 2019 was a year when bloom activity was generally low across this northern region of Lake Champlain, the region of the lake where blooms are typically most common and severe relative to other segments of the lake.

The bloom trends in the northernmost region of the Northeast Arm appear to follow the trends in the southern region of Missisquoi Bay. Both regions had the highest values in 2016 and high values in 2018 and 2020. Missisquoi Bay does have higher maximum individual values in 2018 whereas the Northeast Arm has higher maximum individual values in 2020, but Missisquoi Bay has higher summed values in 2020 than in 2018, which may be why this occurred. Though it may be observed that the Northeast Arm and Missisquoi Bay have similar bloom patterns on a

yearly basis, the data do not present a similarity in pattern across two month periods (Figure 8A, Figure 9A). This could be due to lack of data in the Northeast Arm. However, it does seem that most of the summed cyanobacteria index values for Missisquoi Bay in the two month periods aside from August through September are closer to the eastern and northern borders of the segment, and not in the southern region of the segment that connects to the Northeast Arm. Bloom activity seems to be present in this southern region of Missisquoi Bay mostly in August through September, when bloom activity everywhere is the most severe. So it could be that the bloom activity in the Northeast Arm is mostly influenced by the bloom activity in Missisquoi Bay when it reaches the southernmost region of Missisquoi Bay. In addition, the bloom typically initiates in the northern quadrant of Missisquoi Bay, further suggesting that it may take a while for the bloom to establish connectivity to the Northeast Arm.

Based on Figure 5A, Figure 6A, Figure 1A, and Figure 2A, it is unclear if the eastern region of the Northeast Arm tends to follow the bloom trends in St. Albans Bay. The higher consistency in valid pixels in 2016 would indicate that this is occurring because of high values in St. Albans Bay that spread south from its northern region. However, in 2020 when the Northeast Arm experienced the highest amount of bloom activity throughout the segment, including in the eastern region connecting to St. Albans Bay, St. Albans Bay appeared to have one of its least active years. However, 2020 and 2021 are the only years that St. Albans Bay had valid data spread north to south in the period of June through July, which may contribute to why these years have a higher spread of data in the Northeast Arm, though this cannot be determined.

In St. Albans Bay, although the two shortest blooms had the same duration, one had the latest start date but a typical end date and the smallest maximum spatial extent, while the other had a typical start date but the earliest end date and the highest maximum spatial extent. In

Missisquoi Bay, the two longest blooms did not follow the same pattern of general start date and end date. One of them started earlier, and one of them ended later. These blooms were the third and fourth highest maximum spatial extent and the fourth and third smallest maximum spatial extent. Thus they had typical maximum spatial extent. For these two segments, there appeared to be no obvious pattern between duration, spatial extent, and start and end dates.

In the Northeast Arm, the two longest blooms have relatively typical start dates and late end dates and the two shortest blooms have relatively typical start dates and the two earliest end dates. Although there is not enough data to determine if this is a pattern, but sensor data does suggest that in Missisquoi Bay, the start date is far more consistent (late July/early August), whereas the end date ranges from early September to early November.. These two longest blooms also have the largest maximum spatial extents, although the pattern of highest duration and highest maximum spatial extents does not continue after these two blooms. The third longest bloom has the smallest maximum spatial extent.

In 2016, St. Albans Bay and Missisquoi Bay experienced their highest maximum spatial extents, while the Northeast Arm experienced its second highest (after its 2020 and bloom). In 2019, St. Albans Bay and Missisquoi Bay experienced their lowest maximum spatial extents while the Northeast Arm experienced its second lowest (after 2021). Although in both years an extreme maximum spatial extent did not occur for the Northeast Arm, 2016 was the longest bloom and 2019 was the shortest.

It appears that the reason the Northeast Arm experienced some longer blooms than St. Albans Bay was because of the influence of bloom activity from Missisquoi Bay, which saw the highest duration blooms (Figure 6). It also seems likely that the reason the Northeast Arm typically had such smaller maximum spatial extent is because the bloom activity in this segment

is generally limited to the region directly adjacent to Missisquoi Bay and St. Albans Bay (Figure 7).

It appears that generally, the peak bloom activity begins around half a month earlier in Missisquoi Bay than it does in the Northeast Arm. It also appears that generally the peak bloom activity in St. Albans Bay occurs approximately 2-4 weeks before it occurs in the Northeast Arm. Overall, it seems that the most severe bloom activity in all three segments does occur in August and September suggesting similar summer conditions promote maximum bloom severity.

Tables 6 and 7 provide further evidence that the Northeast Arm is far more significantly affected by the bloom activity in Missisquoi Bay than in St. Albans Bay, in that the proportion of blooms shared by Missisquoi Bay and the Northeast Arm is .2755 higher than the proportion of blooms shared by St. Albans Bay and the Northeast Arm. This difference in proportion is larger than the proportion of blooms shared by all 3 segments, indicating its significance.

Conclusion

From this research, it was found that the three segments of Lake Champlain, Missisquoi Bay, St. Albans Bay, and the Northeast Arm have similar annual trends, though seasonally, Missisquoi Bay and the Northeast Arm behave more similarly than St. Albans Bay. The remotely sensed data of the whole lake over multiple years allowed for this comparison across segments. Though there was significant missing data, in the segments of the lake that had high data usability, the results were consistent with those from other studies, suggesting that using satellite imagery to study blooms in Lake Champlain is a reliable method. This method provides further insight than traditional monitoring by identifying spatial patterns of blooms and collecting the data is cheaper than collecting in-situ data. The process developed for defining and detecting blooms was successful for tracking blooms in the three lake segments that were

highlighted in this research. It worked well in eutrophic segments of the lake for characterizing duration, severity, and distribution of blooms. The methods used did not work well for shorter and smaller blooms that are known to occur in segments of the lake not focused on in this research. To further investigate seasonal trends, more data or developments on certain processes is needed. Though it is unclear how the potential relationship between invalid data and data determined to be not be a bloom, future work could combine these satellite-derived datasets with other environmental data to further study drivers of bloom observations derived from these data.

References

Algae, Phytoplankton and Chlorophyll. (n.d.). Environmental Measurement Systems.

Retrieved September 27, 2021, from

<https://www.fondriest.com/environmental-measurements/parameters/water-quality/algae-phytoplankton-chlorophyll/>

Anderson, D. M., Glibert, P. M., & Burkholder, J. M. (2002). Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries*, 25(4), 704–726. <https://doi.org/10.1007/BF02804901>

Isenstein, E. M., & Park, M.-H. (2014). Assessment of nutrient distributions in Lake Champlain using satellite remote sensing. *Journal of Environmental Sciences (China)*, 26(9), 1831–1836. <https://doi.org/10.1016/j.jes.2014.06.019>

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. *Journal of Great Lakes Research*, 44(3), 436–446.

<https://doi.org/10.1016/j.jglr.2018.04.001>

Bowling, L. C., Blais, S., & Sinotte, M. (2015). Heterogeneous spatial and temporal cyanobacterial distributions in Missisquoi Bay, Lake Champlain: An analysis of a 9 year data set. *Journal of Great Lakes Research*, 41(1), 164–179.

<https://doi.org/10.1016/j.jglr.2014.12.012>

Carlson, R. E. (1977). A Trophic State Index for Lakes. *Limnology and Oceanography*, 22(2), 361–369.

- Carmichael, W. W. (2001). Health Effects of Toxin-Producing Cyanobacteria: “The CyanoHABs.” *Human and Ecological Risk Assessment: An International Journal*, 7(5), 1393–1407. <https://doi.org/10.1080/20018091095087>
- Carmichael, W. W., Azevedo, S. M., An, J. S., Molica, R. J., Jochimsen, E. M., Lau, S., Rinehart, K. L., Shaw, G. R., & Eaglesham, G. K. (2001). Human fatalities from cyanobacteria: Chemical and biological evidence for cyanotoxins. *Environmental Health Perspectives*, 109(7), 663–668.
- Carmichael, W. W., & Boyer, G. L. (2016). Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful Algae*, 54, 194–212. <https://doi.org/10.1016/j.hal.2016.02.002>
- Chen, J., & Quan, W. (2012). Using Landsat/TM Imagery to Estimate Nitrogen and Phosphorus Concentration in Taihu Lake, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1), 273–280. <https://doi.org/10.1109/JSTARS.2011.2174339>
- Chorus, I., & Bartram, J. (Eds.). (1999). *Toxic cyanobacteria in water: A guide to their public health consequences, monitoring, and management*. E & FN Spon.
- Clark, J. M., Schaeffer, B. A., Darling, J. A., Urquhart, E. A., Johnston, J. M., Ignatius, A. R., Myer, M. H., Loftin, K. A., Werdell, P. J., & Stumpf, R. P. (2017). Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. *Ecological Indicators*, 80, 84–95. <https://doi.org/10.1016/j.ecolind.2017.04.046>

Copernicus Sentinel-3 Ocean and Land Colour Instrument (OLCI)—LAADS DAAC. (2018).

Retrieved September 27, 2021, from

<https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/olci/>

Coskun, H. G., Tanik, A., Alganci, U., & Cigizoglu, H. (2008). Determination of Environmental Quality of a Drinking Water Reservoir by Remote Sensing, GIS and Regression Analysis. *Water, Air, and Soil Pollution*, *194*, 275–285.

<https://doi.org/10.1007/s11270-008-9716-x>

Cullen, J. J., Ciotti, Á. M., Davis, R. F., & Lewis, M. R. (1997). Optical detection and assessment of algal blooms. *Limnology and Oceanography*, *42*(5part2), 1223–1239.

https://doi.org/10.4319/lo.1997.42.5_part_2.1223

Dörnhöfer, K., & Opelet, N. (2016). Remote sensing for lake research and monitoring – Recent advances. *Ecological Indicators*, *64*, 105–122.

<https://doi.org/10.1016/j.ecolind.2015.12.009>

Gómez, D., Salvador, P., Sanz, J., & Casanova, J. L. (2021). A new approach to monitor water quality in the Menor sea (Spain) using satellite data and machine learning methods. *Environmental Pollution*, *286*, 117489.

<https://doi.org/10.1016/j.envpol.2021.117489>

Huang, J., Zhang, Y., Arhonditsis, G. B., Gao, J., Chen, Q., & Peng, J. (2020). The magnitude and drivers of harmful algal blooms in China's lakes and reservoirs: A national-scale characterization. *Water Research*, *181*, 115902.

<https://doi.org/10.1016/j.watres.2020.115902>

- Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471–483. <https://doi.org/10.1038/s41579-018-0040-1>
- Hunter, P. D., Tyler, A. N., Gilvear, D. J., & Willby, N. J. (2009). Using Remote Sensing to Aid the Assessment of Human Health Risks from Blooms of Potentially Toxic Cyanobacteria. *Environmental Science & Technology*, *43*(7), 2627–2633. <https://doi.org/10.1021/es802977u>
- Isles, P. D. F., Giles, C. D., Gearhart, T. A., Xu, Y., Druschel, G. K., & Schroth, A. W. (2015). Dynamic internal drivers of a historically severe cyanobacteria bloom in Lake Champlain revealed through comprehensive monitoring. *Journal of Great Lakes Research*, *41*(3), 818–829. <https://doi.org/10.1016/j.jglr.2015.06.006>
- Isles, P. D. F., Xu, Y., Stockwell, J. D., & Schroth, A. W. (2017). Climate-driven changes in energy and mass inputs systematically alter nutrient concentration and stoichiometry in deep and shallow regions of Lake Champlain. *Biogeochemistry*, *133*(2), 201–217. <https://doi.org/10.1007/s10533-017-0327-8>
- Jia, T., Zhang, X., & Dong, R. (2019). Long-Term Spatial and Temporal Monitoring of Cyanobacteria Blooms Using MODIS on Google Earth Engine: A Case Study in Taihu Lake. *Remote Sensing*, *11*(19), 2269. <https://doi.org/10.3390/rs11192269>
- Kim, H.-C., Son, S., Kim, Y. H., Khim, J. S., Nam, J., Chang, W. K., Lee, J.-H., Lee, C.-H., & Ryu, J. (2017). Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. *Marine Pollution Bulletin*, *121*(1), 425–434. <https://doi.org/10.1016/j.marpolbul.2017.05.026>

Lake and Basin Facts. (n.d.). Lake Champlain Basin Program. Retrieved September 27, 2021, from <https://www.lcbp.org/about-the-basin/facts/>

Lake Champlain Lay Monitoring Program Report 1979 – 2013. (2013). Retrieved September 30, 2021, from https://anrweb.vt.gov/PubDocs/DEC/WSMD/Lakes/Docs/lp_2013Impreport-champlain.pdf

Lake Look: Five Lakes in One. (2020). Lake Champlain Committee. Retrieved September 27, 2021, from

<https://www.lakechamplaincommittee.org/learn/news/item/lake-look-five-lakes-in-one>

Levine, S. N., Lini, A., Ostrofsky, M. L., Bunting, L., Burgess, H., Leavitt, P. R., Reuter, D., Lami, A., Guilizzoni, P., & Gilles, E. (2012). The eutrophication of Lake Champlain's northeastern arm: Insights from paleolimnological analyses. *Journal of Great Lakes Research*, 38, 35–48. <https://doi.org/10.1016/j.jglr.2011.07.007>

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 Sensing of Environment*, 157, 24–34. <https://doi.org/10.1016/j.rse.2014.06.008>

Melendez-Pastor, I., Isenstein, E. M., Navarro-Pedreño, J., & Park, M.-H. (2019). Spatial variability and temporal dynamics of cyanobacteria blooms and water quality parameters in Missisquoi Bay (Lake Champlain). *Water Supply*, 19(5), 1500–1506. <https://doi.org/10.2166/ws.2019.017>

- 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. *Scientific Reports*, 9(1), 18310. <https://doi.org/10.1038/s41598-019-54453-y>
- National Centers for Coastal Ocean Sciences. (2022). *Lake Champlain 300m* [Data set]. https://products.coastalscience.noaa.gov/habs_explorer/index.php?path=djFYWEE0NURTMllpTkN5VUVzdmtlbGxQalBCYVd5REsvL0lVYzNJaVF2NIY1a1F2TS9tbIJyWjZxdnV4NVE3eQ==&uri=VWtuM1UzbVNVN0RsZzJMeTJvNINpM29OalF0WTFQQjVZVnpuS3o5bnh1Ym0vYWhtWEh4ck1hREVUamE4SDZ0M0JnSnNMaHk4U2YyaTc0R04zM2ZlId0hBbkVjOGkrTnJjMjR3WmtJbGJFeUx4OUVOMlc2TEV4ckdQSXZhQWRZQXQ=&type=blIEUXA3TmhSK21RVDlqbFYxMmEwdz09
- Nutrients and Eutrophication*. (2019). Retrieved September 27, 2021, from https://www.usgs.gov/mission-areas/water-resources/science/nutrients-and-eutrophication?qt-science_center_objects=0#qt-science_center_objects
- Paerl, H., & Huisman, J. (2009). Climate Change: A Catalyst for Global Expansion of Harmful Cyanobacterial Blooms. *Environmental Microbiology Reports*, 1, 27–37. <https://doi.org/10.1111/j.1758-2229.2008.00004.x>
- Paerl, H. W., & Paul, V. J. (2012). Climate change: Links to global expansion of harmful cyanobacteria. *Water Research*, 46(5), 1349–1363. <https://doi.org/10.1016/j.watres.2011.08.002>
- Pahlevan, N., Ackleson, S. G., & Shaeffer, B. A. (2018). Toward a Satellite-Based Monitoring System for Water Quality. *Eos*. Retrieved September 9, 2021, from <https://eos.org/meeting-reports/toward-a-satellite-based-monitoring-system-for-water-quality>

- Pyo, J., Duan, H., Baek, S., Kim, M. S., Jeon, T., Kwon, Y. S., Lee, H., & Cho, K. H. (2019). A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. *Remote Sensing of Environment*, 233, 111350. <https://doi.org/10.1016/j.rse.2019.111350>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Rosen, B. H., Shambaugh, A., Ferber, L., Smith, F., Watzin, M., Eliopoulos, C., & Stangel, P. (2001). *Evaluation of Potential Blue-Green Algal Toxins In Lake Champlain, Summer 2000*. 27.
- Ruddick, K. G., Gons, H. J., Rijkeboer, M., & Tilstone, G. (2001). Optical remote sensing of chlorophyll a in case 2 waters by use of an adaptive two-band algorithm with optimal error properties. *Applied Optics*, 40(21), 3575–3585. <https://doi.org/10.1364/AO.40.003575>
- Schindler, D. W. (2012). The dilemma of controlling cultural eutrophication of lakes. *Proceedings of the Royal Society B: Biological Sciences*, 279(1746), 4322–4333. <https://doi.org/10.1098/rspb.2012.1032>
- Sellner, K. G., Doucette, G. J., & Kirkpatrick, G. J. (2003). Harmful algal blooms: Causes, impacts and detection. *Journal of Industrial Microbiology and Biotechnology*, 30(7), 383–406. <https://doi.org/10.1007/s10295-003-0074-9>
- Smeltzer, E., Dunlap, F., & Simoneau, M. (2009). *Lake Champlain Phosphorus Concentrations and Loading Rates, 1990-2008*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014).

Dropout: A Simple Way to Prevent Neural Networks from Overfitting. 30.

Stanier, R. Y., & Cohen-Bazire, G. (1977). Phototrophic Prokaryotes: The Cyanobacteria.

Annual Review of Microbiology, 31(1), 225–274.

<https://doi.org/10.1146/annurev.mi.31.100177.001301>

Stewart, I., Webb, P. M., Schluter, P. J., & Shaw, G. R. (2006). Recreational and

occupational field exposure to freshwater cyanobacteria – a review of anecdotal and case reports, epidemiological studies and the challenges for epidemiologic assessment.

Environmental Health, 5, 6. <https://doi.org/10.1186/1476-069X-5-6>

Stumpf, R. P., Davis, T. W., Wynne, T. T., Graham, J. L., Loftin, K. A., Johengen, T. H.,

Gossiaux, D., Palladino, D., & Burtner, A. (2016). Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae*, 54, 160–173.

<https://doi.org/10.1016/j.hal.2016.01.005>

Taranu, Z. E., Gregory-Eaves, I., Leavitt, P. R., Bunting, L., Buchaca, T., Catalan, J.,

Domaizon, I., Guilizzoni, P., Lami, A., McGowan, S., Moorhouse, H., Morabito, G.,

Pick, F. R., Stevenson, M. A., Thompson, P. L., & Vinebrooke, R. D. (2015).

Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the

Anthropocene. *Ecology Letters*, 18(4), 375–384. <https://doi.org/10.1111/ele.12420>

Torbick, N., & Corbiere, M. (2015). A Multiscale Mapping Assessment of Lake Champlain

Cyanobacterial Harmful Algal Blooms. *International Journal of Environmental*

Research and Public Health, 12(9), 11560–11578.

<https://doi.org/10.3390/ijerph120911560>

- Trescott, A., & Park, M.-H. (2013). Remote sensing models using Landsat satellite data to monitor algal blooms in Lake Champlain. *Water Science and Technology*, 67(5), 1113–1120. <https://doi.org/10.2166/wst.2013.661>
- Tyler, A. N., Hunter, P. D., Spyrakos, E., Groom, S., Constantinescu, A. M., & Kitchen, J. (2016). Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. *Science of The Total Environment*, 572, 1307–1321. <https://doi.org/10.1016/j.scitotenv.2016.01.020>
- 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>
- Vander Woude, A., Ruberg, S., Johengen, T., Miller, R., & Stuart, D. (2019). Spatial and temporal scales of variability of cyanobacteria harmful algal blooms from NOAA GLERL airborne hyperspectral imagery. *Journal of Great Lakes Research*, 45(3), 536–546. <https://doi.org/10.1016/j.jglr.2019.02.006>
- Watson, S. B., Miller, C., Arhonditsis, G., Boyer, G. L., Carmichael, W., Charlton, M. N., Confesor, R., Depew, D. C., Höök, T. O., Ludsin, S. A., Matisoff, G., McElmurry, S. P., Murray, M. W., Peter Richards, R., Rao, Y. R., Steffen, M. M., & Wilhelm, S. W. (2016). The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. *Harmful Algae*, 56, 44–66. <https://doi.org/10.1016/j.hal.2016.04.010>
- Watzin, M. C., Miller, E. B., Shambaugh, A. D., & Kreider, M. A. (2006). Application of the WHO alert level framework to cyanobacterial monitoring of Lake Champlain, Vermont. *Environmental Toxicology*, 21(3), 278–288. <https://doi.org/10.1002/tox.20181>

- Wynne, T. T., & Stumpf, R. P. (2015). Spatial and Temporal Patterns in the Seasonal Distribution of Toxic Cyanobacteria in Western Lake Erie from 2002–2014. *Toxins*, 7(5), 1649–1663. <https://doi.org/10.3390/toxins7051649>
- Xu, Y., Schroth, A. W., Isles, P. D. F., & Rizzo, D. M. (2015a). Quantile regression improves models of lake eutrophication with implications for ecosystem-specific management. *Freshwater Biology*, 60(9), 1841–1853. <https://doi.org/10.1111/fwb.12615>
- Xu, Y., Schroth, A. W., & Rizzo, D. M. (2015b). Developing a 21st Century framework for lake-specific eutrophication assessment using quantile regression. *Limnology and Oceanography: Methods*, 13(5), 237–249. <https://doi.org/10.1002/lom3.10021>
- Zamyadi, A., MacLeod, S. L., Fan, Y., McQuaid, N., Dorner, S., Sauvé, S., & Prévost, M. (2012). Toxic cyanobacterial breakthrough and accumulation in a drinking water plant: A monitoring and treatment challenge. *Water Research*, 46(5), 1511–1523. <https://doi.org/10.1016/j.watres.2011.11.012>

Appendix A

Sum and Max Raster Calculations Grouped by Segments

This appendix consists of the summation and maximum calculations for composites grouped and scaled by segment and time period. The white region shown in Figures 10, 11, 14, 15, 16, and 18 represents land in the lake segment and the pixel color that covers this segment in other plots is representative of the water portion of the pixel.

Figure 1A

Cyanobacteria Index summations for composites grouped by year for St. Albans Bay

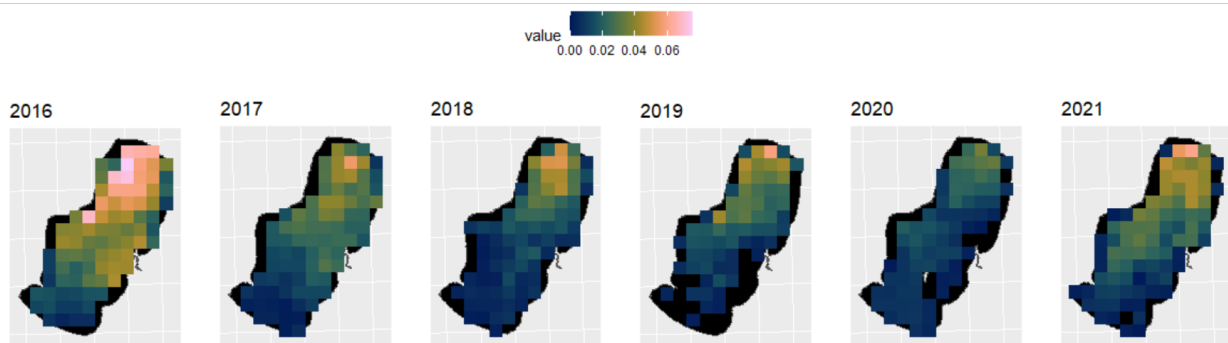


Figure 2A

Cyanobacteria Index maximum calculations for composites grouped by year for St. Albans Bay

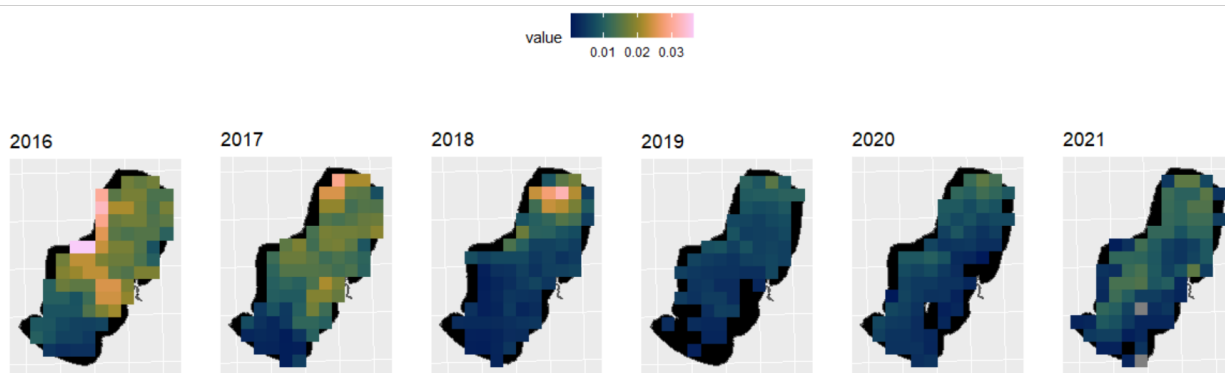


Figure 3A

Cyanobacteria Index summations for composites grouped by year for Missisquoi Bay

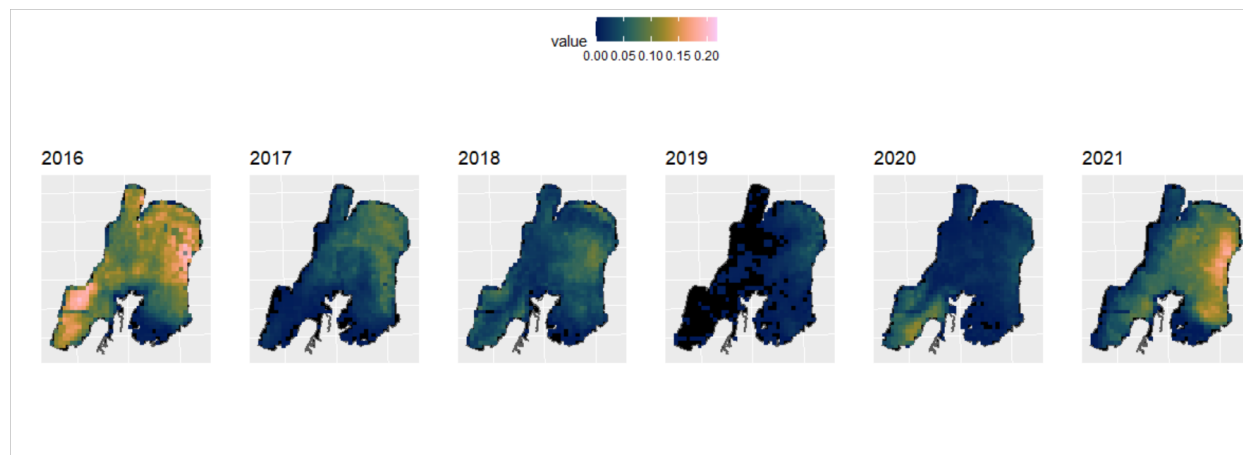


Figure 4A

Cyanobacteria Index maximum calculations for composites grouped by year for Missisquoi Bay

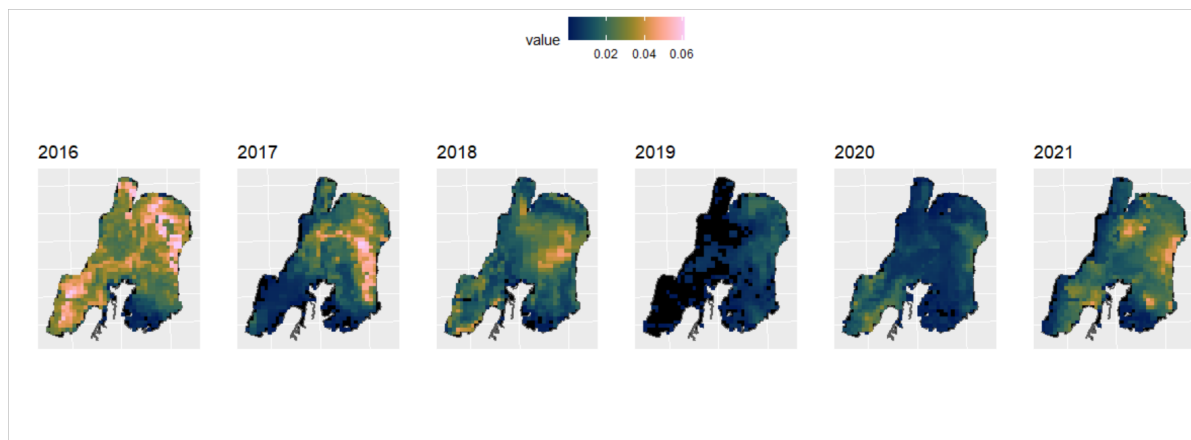


Figure 5A

Cyanobacteria Index summations for composites grouped by year for the Northeast Arm

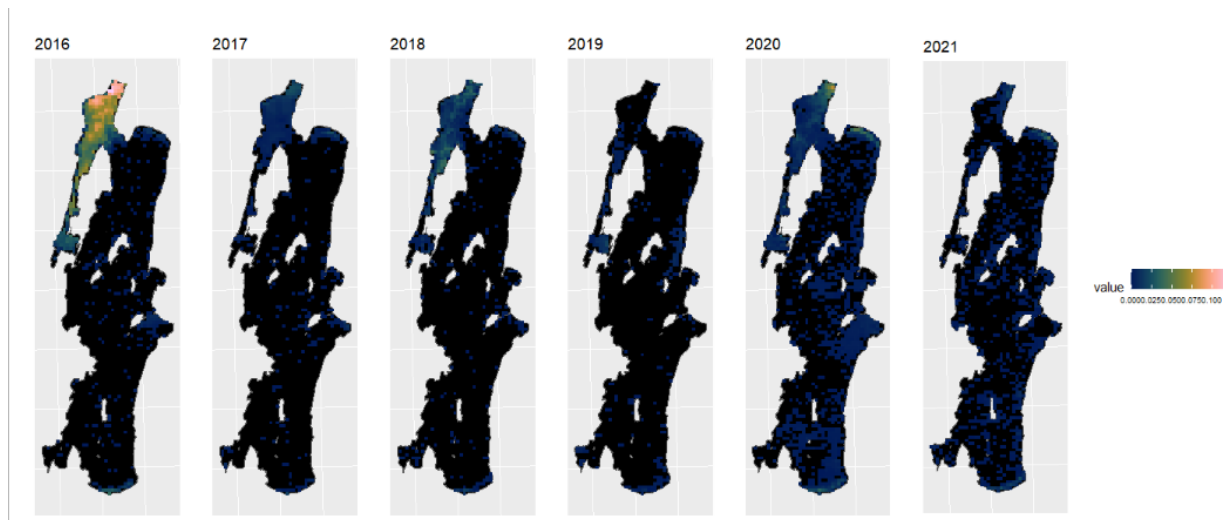


Figure 6A

Cyanobacteria Index maximum calculations for composites grouped by year for the Northeast

Arm

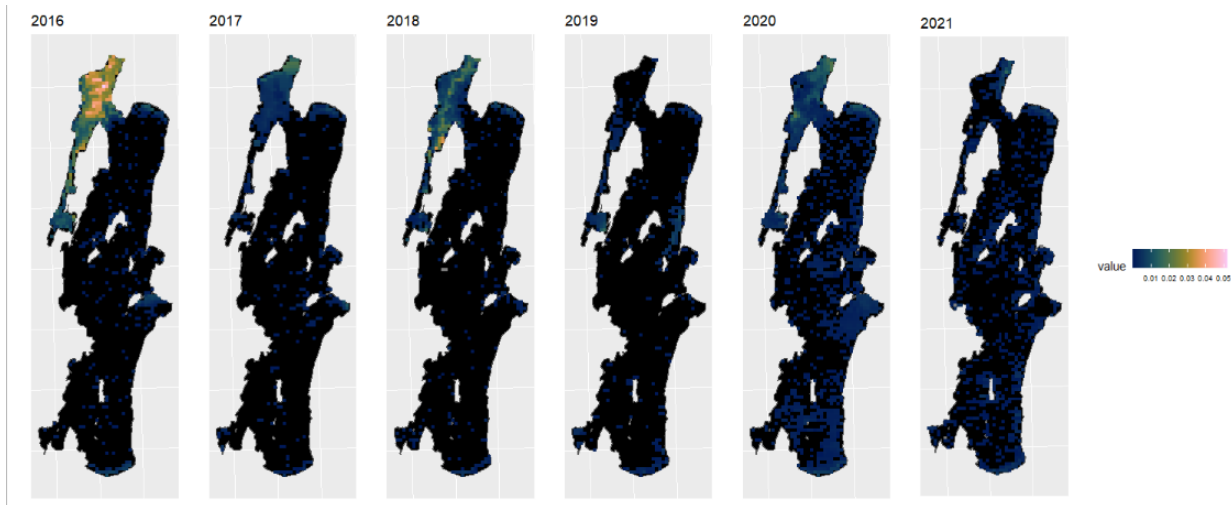


Figure 7A

Cyanobacteria Index summation calculations for composites grouped by 2 month periods between April and November separated by year for St. Albans Bay

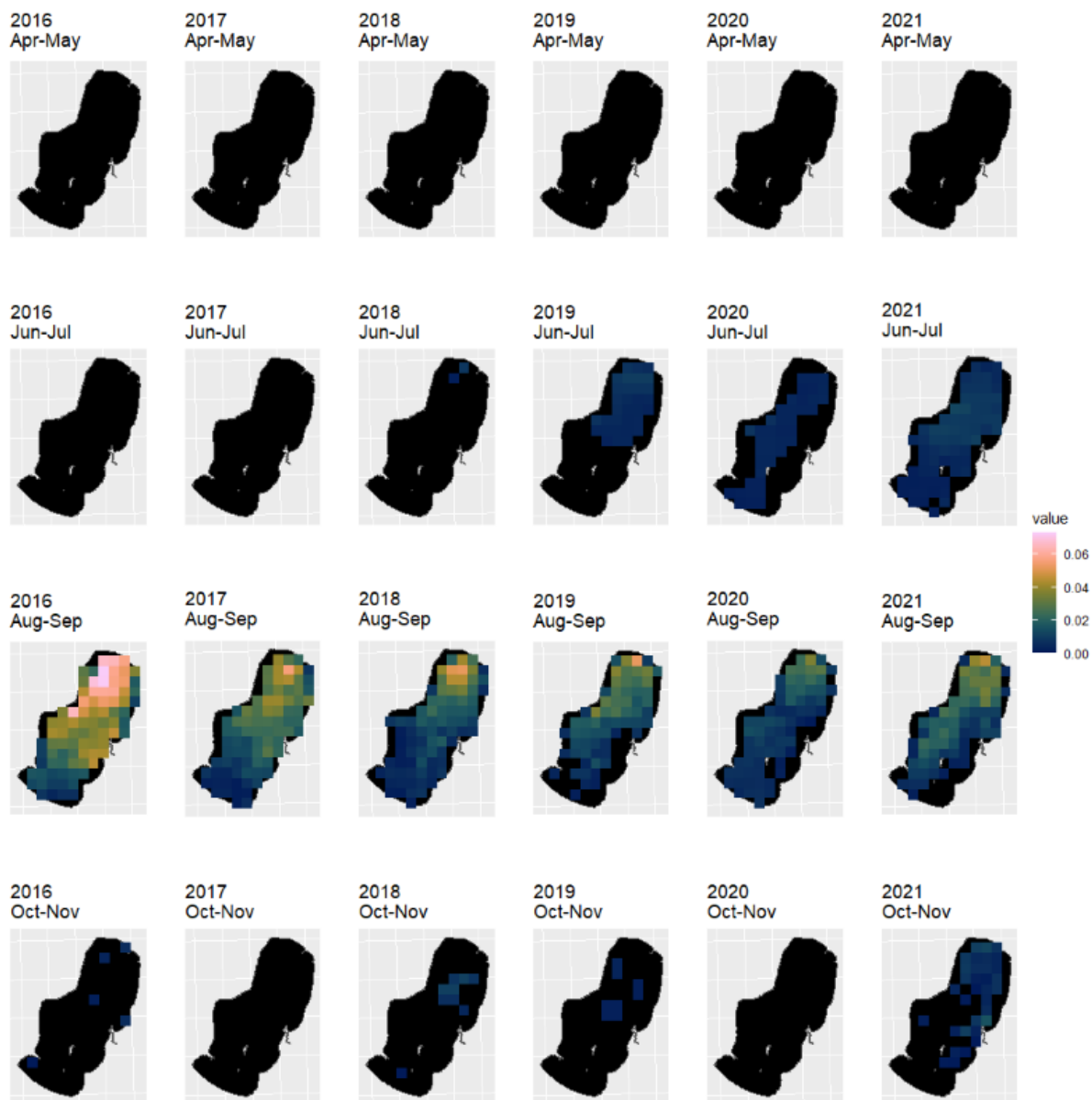


Figure 8A

Cyanobacteria Index summation calculations for composites grouped by 2 month periods between April and November separated by year for Missisquoi Bay

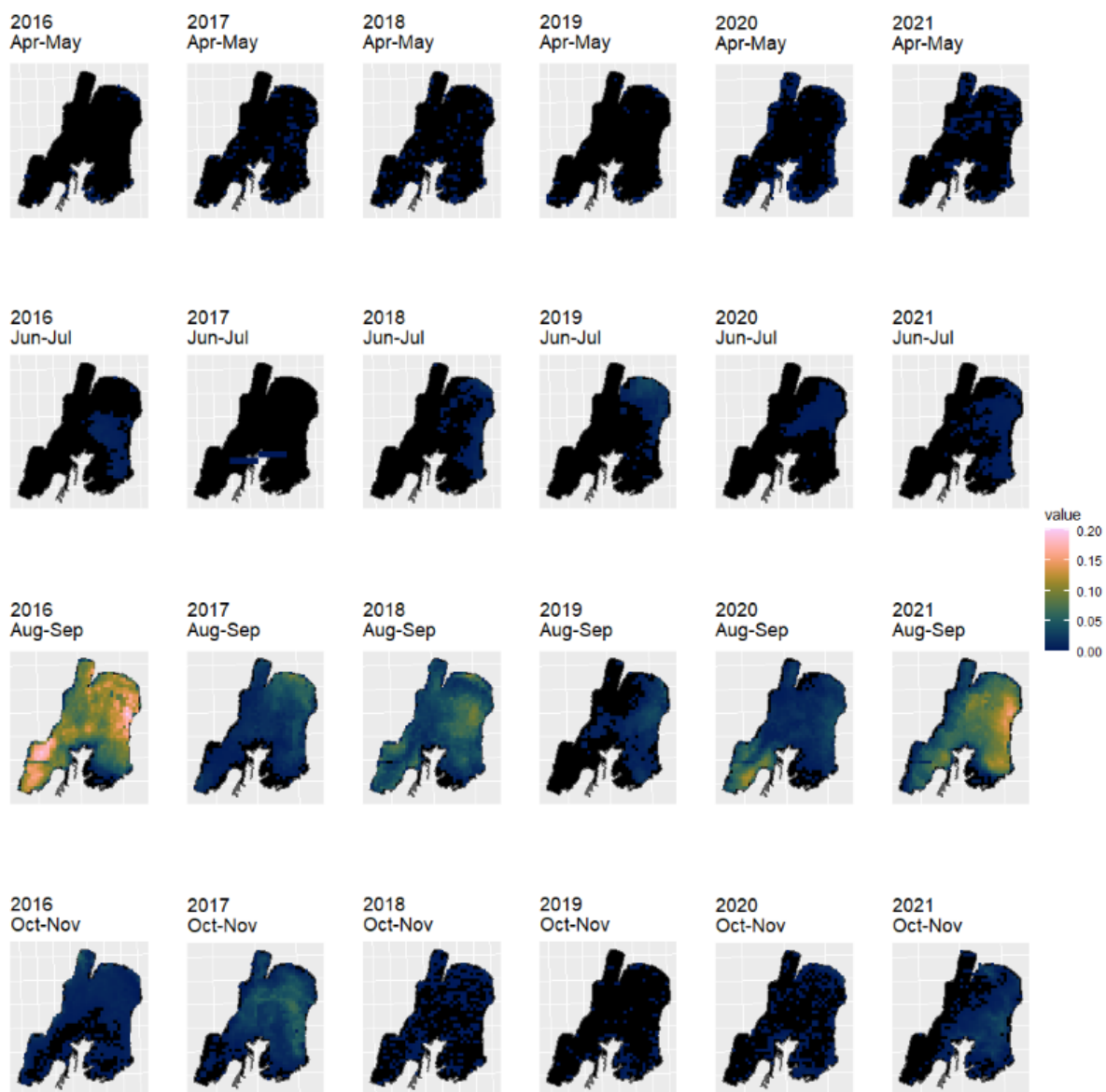
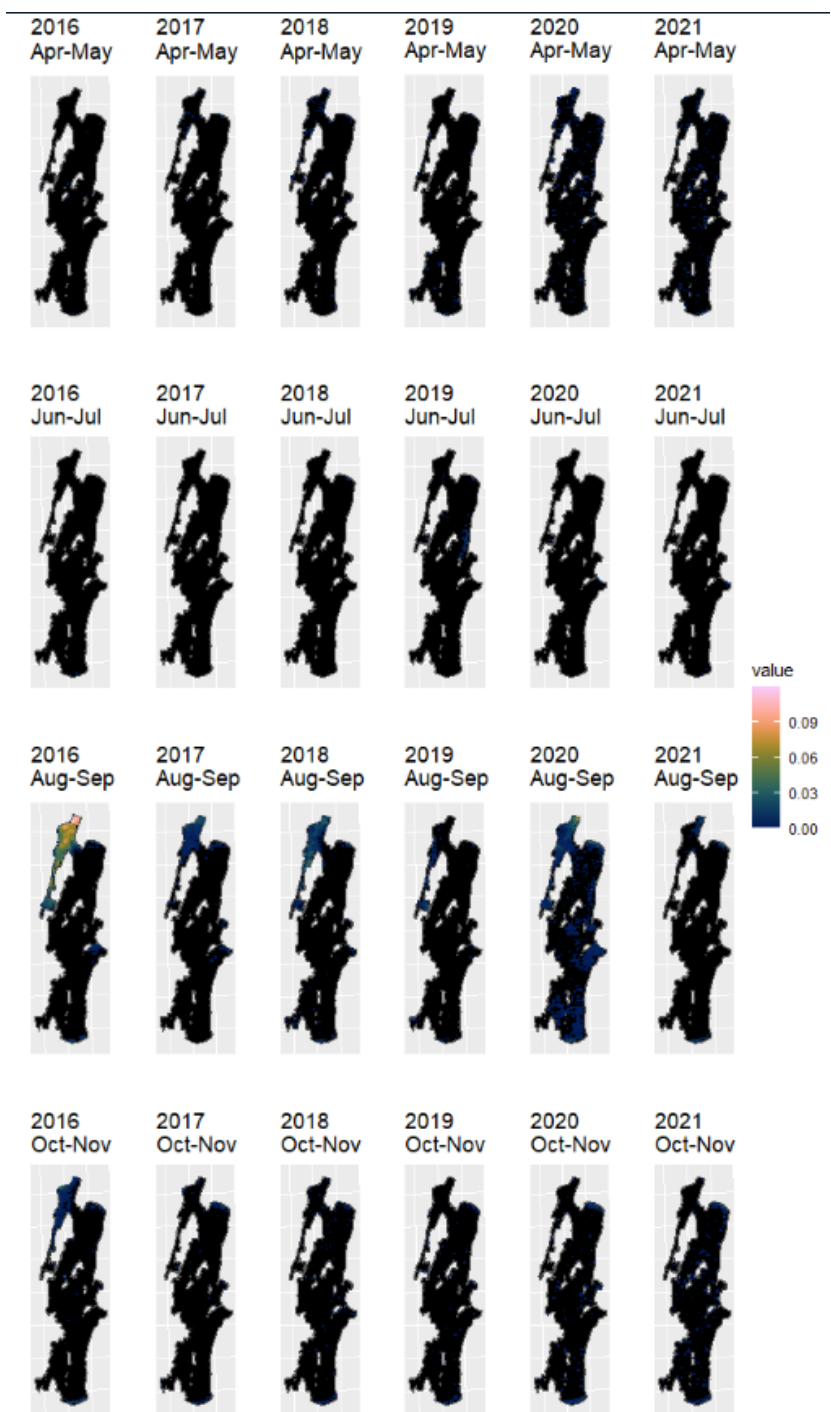


Figure 9A

Cyanobacteria Index summation calculations for composites grouped by 2 month periods

between April and November separated by year for the Northeast Arm



St. Albans Bay 2017 July - November summed CI value total per composite

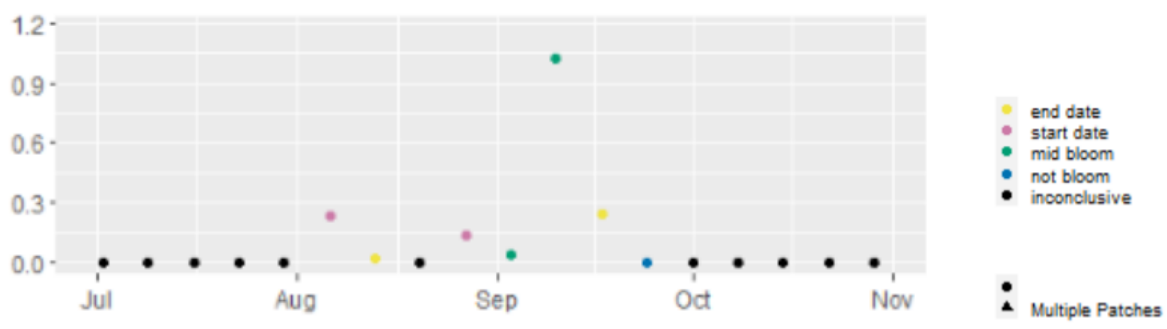


Figure 13B

Northeast Arm 2018 July - November maximum CI value per composite

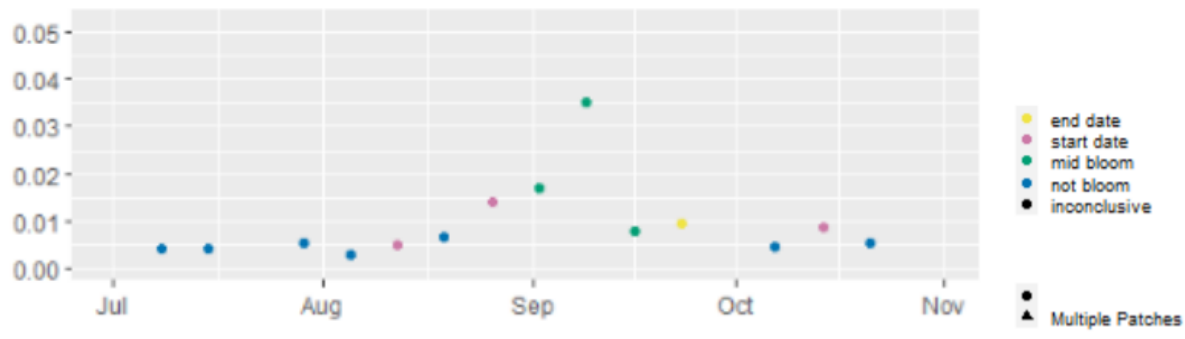


Figure 14B

Northeast Arm 2018 July - November summed CI value total per composite

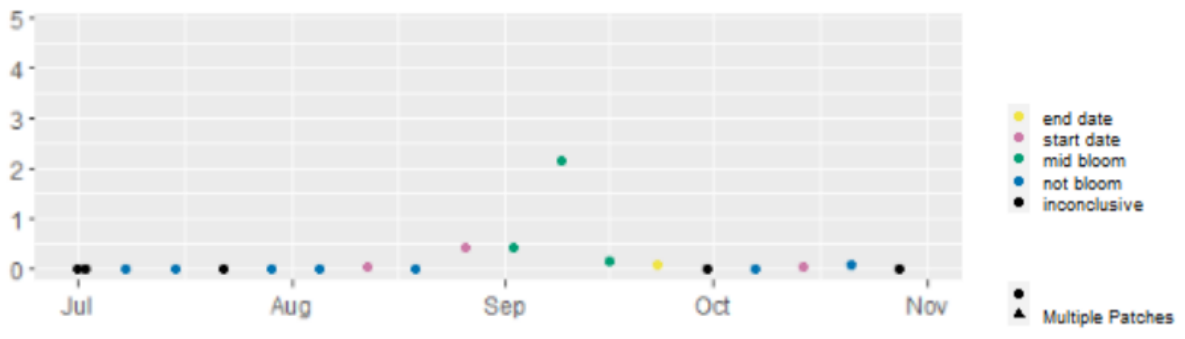


Figure 15B

Missisquoi Bay 2019 July - November maximum CI value per composite

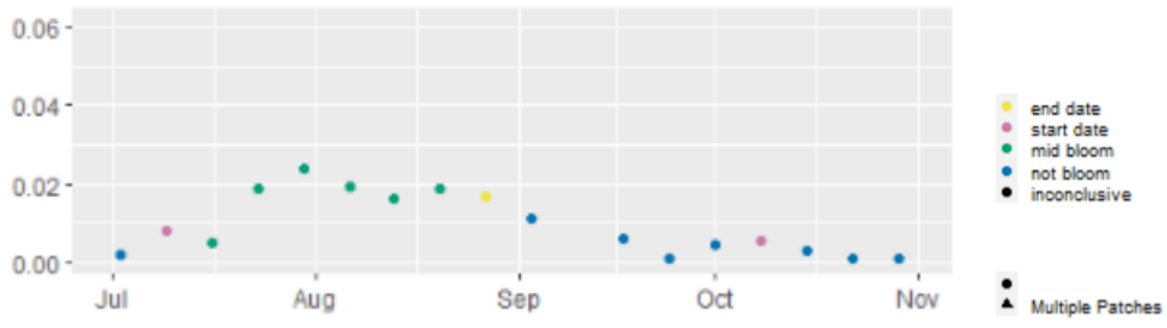


Figure 22B

Missisquoi Bay 2019 July - November summed CI value total per composite

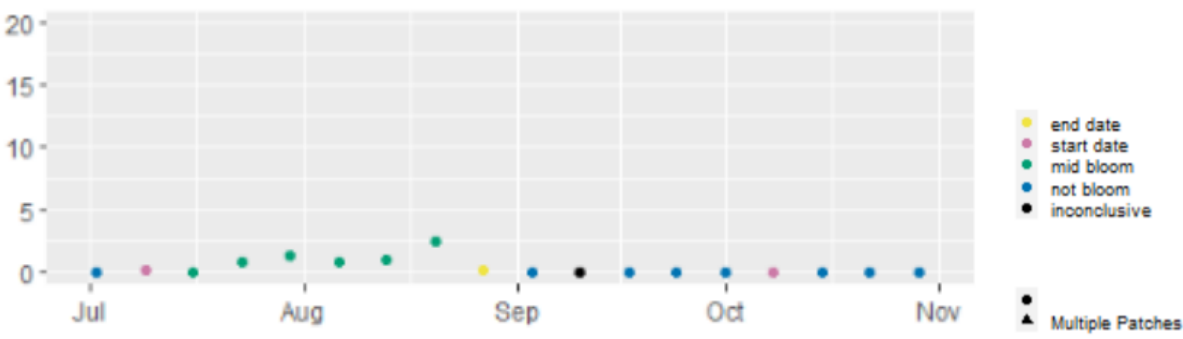


Figure 23B

St. Albans 2019 July - November maximum CI value per composite

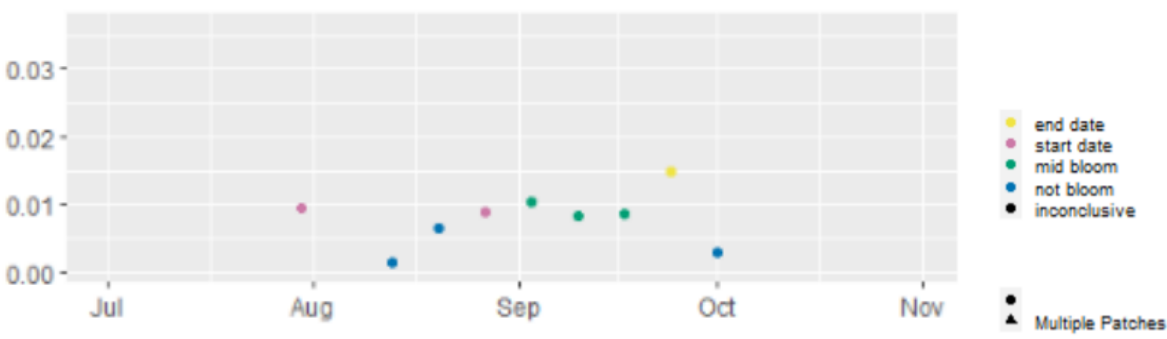


Figure 24B

St. Albans 2019 July - November summed CI value total per composite

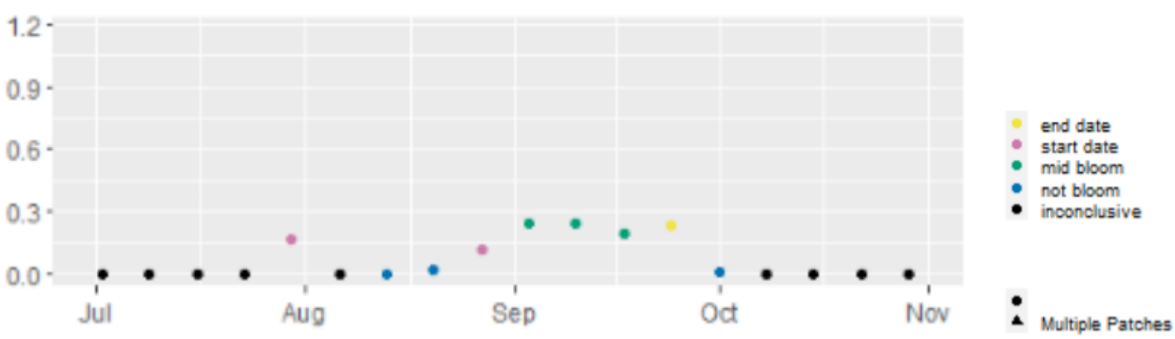


Figure 25B

Northeast Arm 2020 July - November maximum CI value per composite

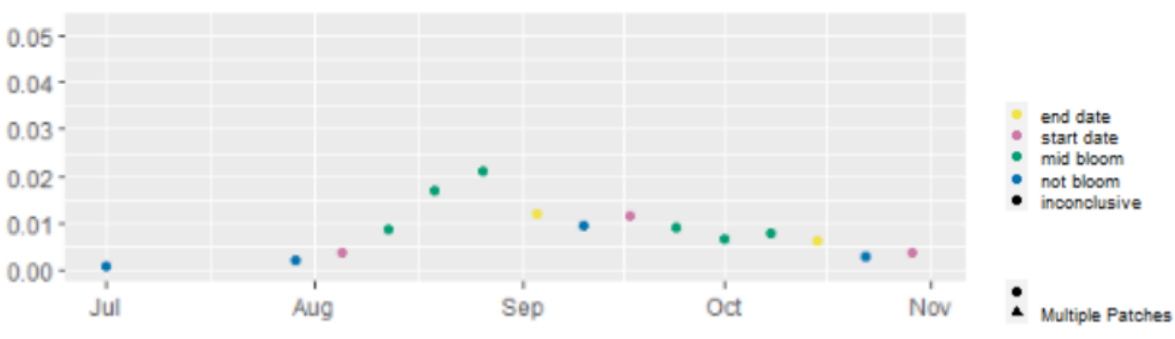


Figure 26B

Northeast Arm 2020 July - November summed CI value total per composite

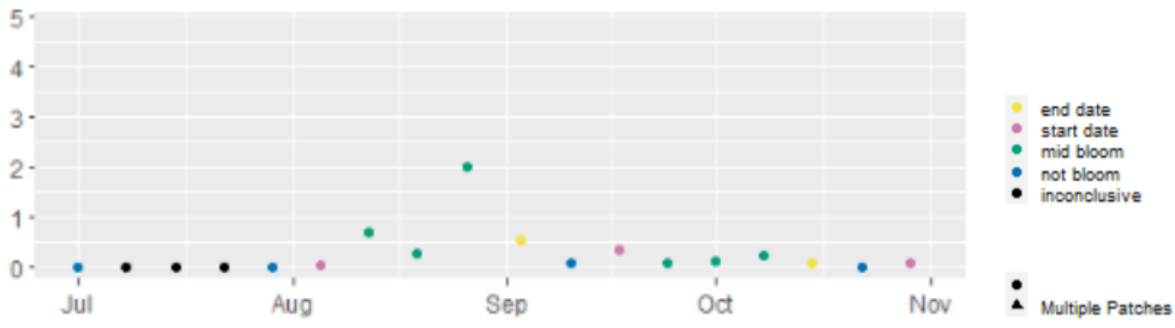


Figure 27B

Missisquoi Bay 2020 July - November maximum CI value per composite

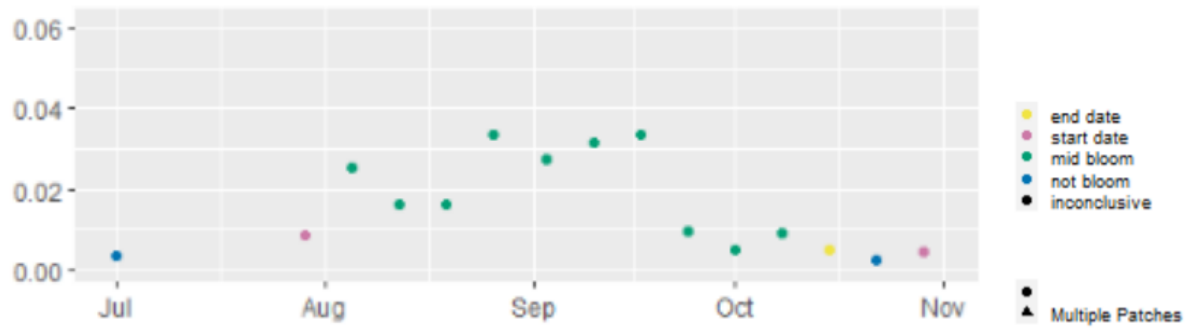


Figure 28B

Missisquoi Bay 2020 July - November summed CI value total per composite

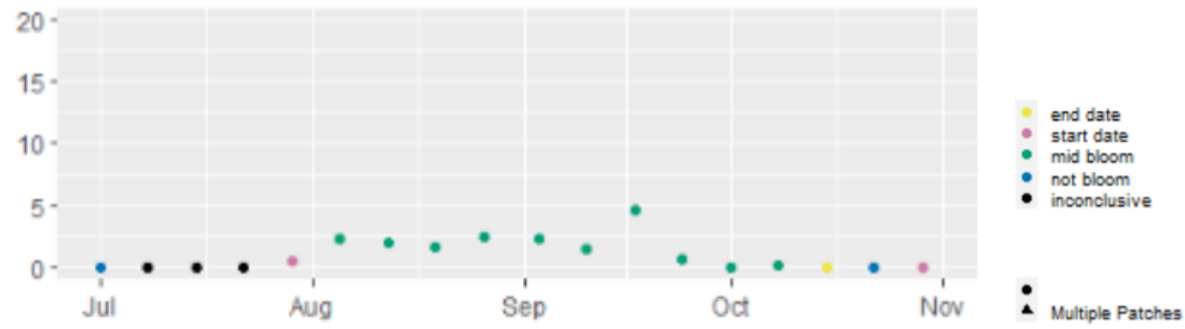


Figure 29B

St. Albans Bay 2020 July - November maximum CI value per composite

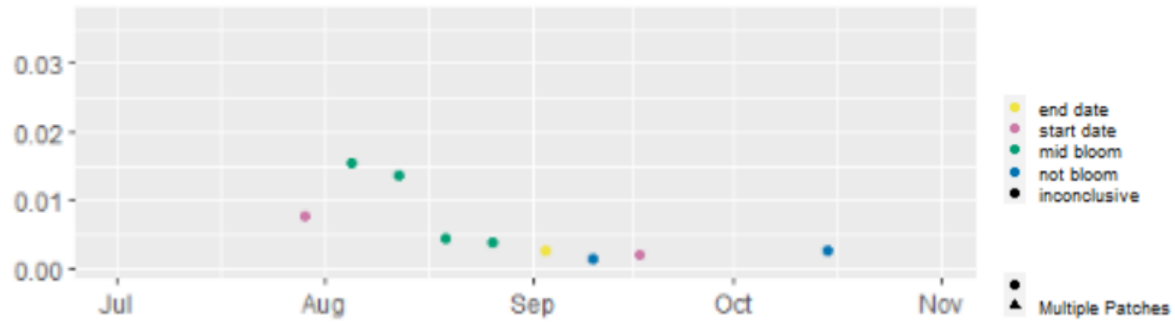


Figure 30B

Missisquoi Bay 2021 July - November maximum CI value per composite

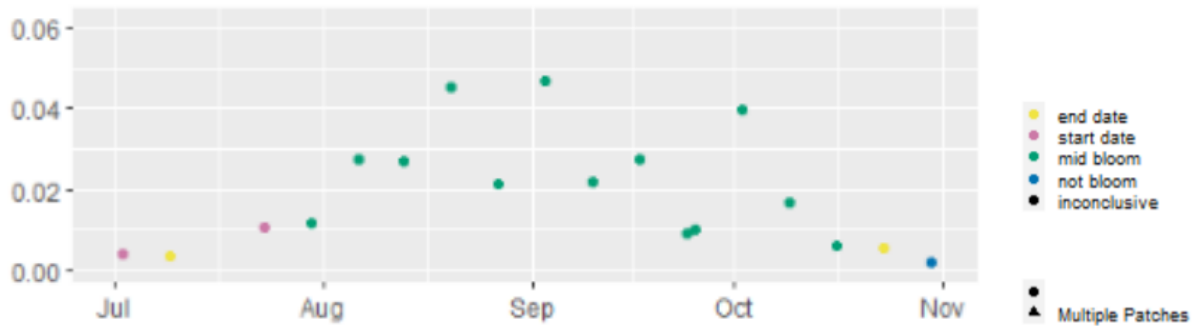


Figure 34B

Missisquoi Bay 2021 July - November summed CI value total per composite

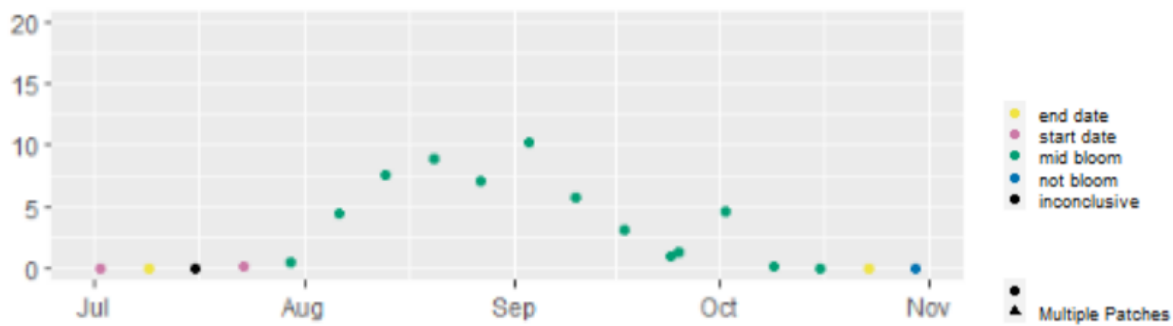


Figure 35B

St. Albans Bay 2021 July - November maximum CI value per composite

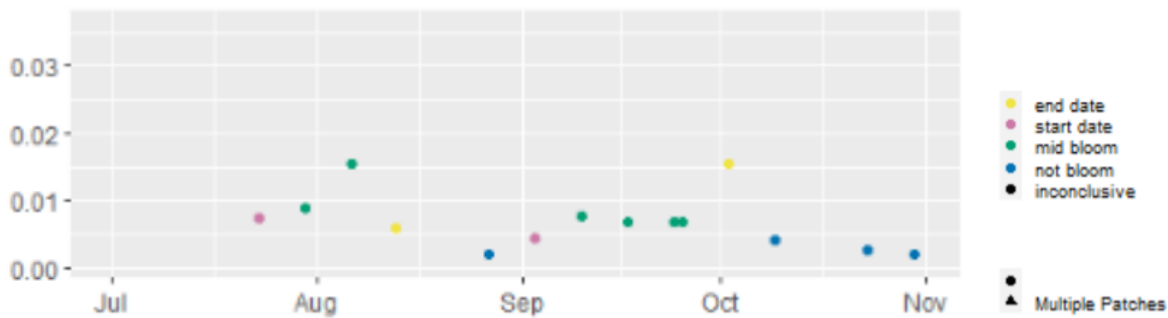


Figure 36B

St. Albans Bay 2021 July - November summed CI value total per composite

