



Brigham Young University
BYU ScholarsArchive

Theses and Dissertations

2025-12-16

Utah Lake From Space: Analysis of Sentinel 2 and MODIS Data Shows Spatiotemporal Trends and Patterns in Chlorophyll-a, Turbidity, and Temperature

Kaylee Brook Tanner
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

Part of the Engineering Commons

BYU ScholarsArchive Citation

Tanner, Kaylee Brook, "Utah Lake From Space: Analysis of Sentinel 2 and MODIS Data Shows Spatiotemporal Trends and Patterns in Chlorophyll-a, Turbidity, and Temperature" (2025). *Theses and Dissertations*. 11093.

<https://scholarsarchive.byu.edu/etd/11093>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Utah Lake From Space: Analysis of Sentinel 2 and MODIS Data

Shows Spatiotemporal Trends and Patterns in

Chlorophyll-a, Turbidity, and Temperature

Kaylee Brook Tanner

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Gustavious Williams, Chair
Woodruff Miller
Robert Sowby
Candace Berrett

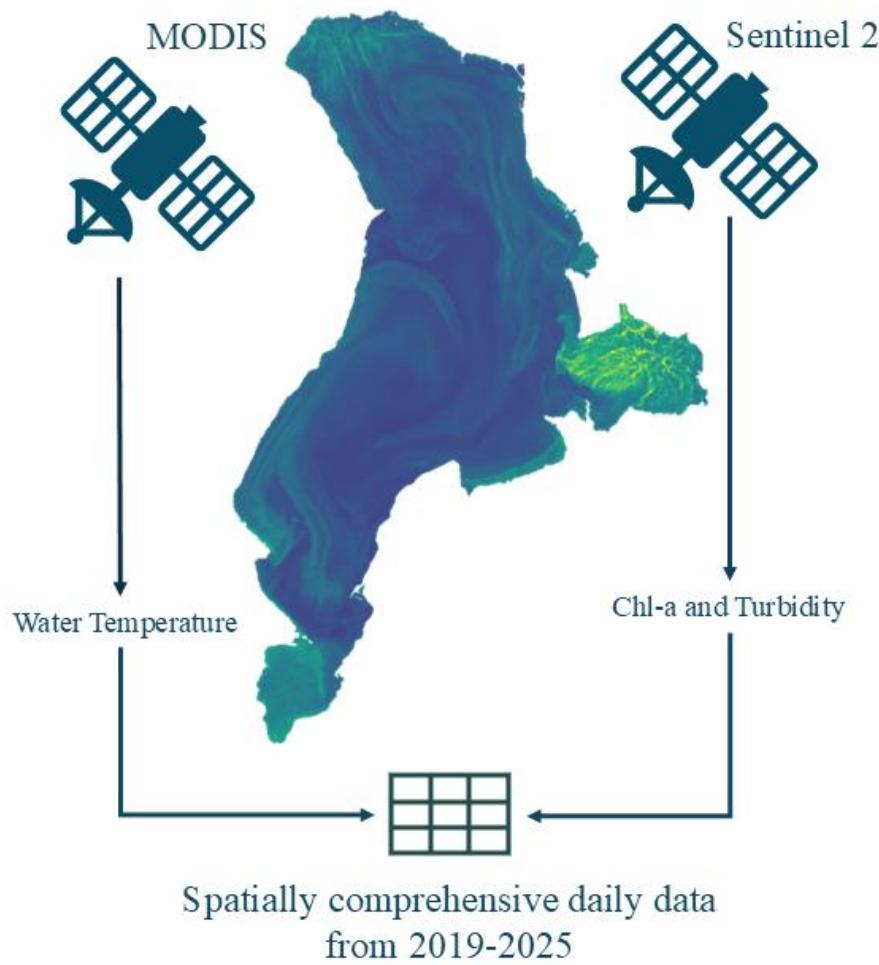
Department of Civil and Construction Engineering

Brigham Young University

Copyright © 2025 Kaylee Brook Tanner

All Rights Reserved

Utah Lake From Space: Analysis of Sentinel 2 and MODIS Data Shows Spatiotemporal Trends and Patterns in Chlorophyll-a, Turbidity, and Temperature



Kaylee Brook Tanner

Department of Civil and Construction Engineering
Doctor of Philosophy

BYU ENGINEERING

Abstract

Harmful algal blooms (HABs) are a serious and growing risk to waterbodies around the world. Excessive algal growth, especially when it involves cyanobacteria, impairs the functioning of aquatic ecosystems and threatens the health of humans and animals using the water. There is currently great interest in developing effective prevention and mitigation strategies for these blooms, particularly in ecologically and economically vital freshwater lakes. Such development, however, requires detailed understanding of HAB drivers and dynamics, which can be highly complex and difficult to characterize. High-frequency, high-resolution imagery from earth observation satellites provides spatially and temporally comprehensive data on global lakes—data which can be used to fill gaps in understanding left by in situ studies of HABs, which are by nature limited in spatial and temporal scope.

We developed a workflow for extracting accurate, usable estimates of chlorophyll-a (chl-a, a plant pigment commonly used as an index for algal biomass), turbidity, and water temperature from the ESA's Sentinel 2 satellite series and NASA's MODIS sensor using the Google Earth Engine (GEE) platform. This workflow includes a novel approach to generating chl-a and turbidity retrieval algorithms, combining physics-based and empirical models to generate more accurate and explainable estimates of both chl-a and turbidity measurements, as well as methods for QA/QC and subsampling of imagery.

This workflow can be easily adapted to any lake around the world, but we demonstrate it on a case study of Utah Lake, a large, shallow, eutrophic lake and vitally important natural resource in the semi-arid Utah Valley, Utah, USA. We use the dataset obtained from the GEE workflow and some additional processed Sentinel 2

imagery to analyze large-scale patterns in algal blooms on Utah Lake and the relationships between algal blooms and water column temperature and turbidity.

Our results confirm pronounced spatial heterogeneity in the occurrence of algal blooms on Utah Lake and reveal distinct intra- and inter-seasonal trends. The shallow, hydrologically isolated Provo Bay on the east side of the lake was identified as a major bloom hotspot. Another bay at the south end of the lake, Goshen Bay, was also a hotspot, but to a lesser extent than Provo Bay. Both bays also appear to act as bloom “incubators” which tend to bloom earlier in the season and push water with high chl-a concentrations out into the main lake. We found that intense blooms were relatively rare, especially on the open lake away from the shoreline and the shallow bays, and were moderately associated with lower turbidity values.

We use these results to make several recommendations for future bloom monitoring and mitigation strategies on Utah Lake. Specifically, our findings suggest that wetland restorations and hydrologic management may be more effective than reducing specific nutrient inflows, and that monitoring and HAB advisory efforts should account for the spatial variability of blooms. This research provides valuable insight for policymakers, water quality managers, and others involved in Utah Lake research and management. In addition, it supports ongoing development of best practices and methods for using satellite imagery to study water quality, and provides an accessible tool which can be applied to any waterbody in the world to conduct a similar analysis.

Keywords: Utah Lake, chl-a, remote sensing, Sentinel, MODIS, algal bloom

Acknowledgments

This work represents great investments in me that have been made by my advisor, Dr. Gustavious Williams, as well as by Brigham Young University, the Department of Civil and Engineering, and my family and friends. I am profoundly grateful for the support I have received. I don't know that everyone can say graduate school was the happiest and most fulfilling time of their life so far. I feel grateful that this was the case for me and want to express my deepest appreciation for the people and institutions that made this possible.

In addition to generous financial support from the Department of Civil Engineering in the form of scholarships and teaching assistantships, the Utah NASA Space Grant Consortium funded a large portion of this research, and the rest was covered by various research grants through Dr. Williams. This enabled me to focus completely on my studies and research for almost the entirety of my program, which I greatly appreciate.

I am grateful for my husband, who perhaps did not fully understand what he was getting himself into when he married a 4th-year PhD student but nevertheless did all of the cooking and housework for weeks at a time without complaint while I worked. He is wonderfully supportive and I love and admire him with all my heart.

I am grateful for my family and friends, who never stopped encouraging me and teasing me with "Doctor Kaylee" even when I didn't manage to graduate on the timeline I had originally planned. I am especially thankful to my parents, who instilled in me a deep belief in my own capacity for challenging work, set an example of hard work and chasing lofty goals, and supported me financially and emotionally throughout my university career.

I have no words to express how honored and grateful I am to have worked with Dr. Williams and my main student collaborator on this project, Anna Cardall, for the last several years. Anna is not only the best coworker I've ever had, but also became the best friend I've ever had. Not having a desk next to hers is probably the thing I will miss most about graduate school. Dr. Williams is an incredibly talented, dedicated, and patient mentor, and I owe not only this work but also a significant part of my character and attributes as a scientist and engineer (as well as my writing style) to his training and influence. I genuinely believe I would not have made it through a PhD with anyone else as my mentor.

Finally, I am grateful to my Father in Heaven—for guiding my path in the most unexpected but wonderful ways, for filling my life with the best people in the world, and for giving me the capacity and opportunity to develop skills that I can use in service of humanity's sacred stewardship over the Earth.

Table of Contents

Abstract	iii
Acknowledgments	v
Table of Contents	vii
Chapter 1: Introduction	1
1.1 Utah Lake and Algal Blooms	1
1.3 Remote sensing of algal blooms	5
Chapter 2: A six-year, spatiotemporally comprehensive Utah Lake chl-a dataset	9
2.1 Introduction	9
2.2 Data Description	11
2.3 Methods	12
2.3.1 Sentinel image processing	12
2.3.2 Remote sensing models	13
2.3.3 In situ data	14
2.3.4 Combining physics-based and empirical models	15
2.3.5 Chlorophyll-a	17
3.3.4 Turbidity	18
3.3.5 Temperature	19
2.4 Image sampling and model application	20
2.4.1 Whole-lake samples	21
2.4.2 Boxes sample	23
2.4.3 Machine learning near-shore, open-water clusters	25
2.4.4 Model application	27
2.4 Imputation of missing and anomalous values	27
2.4.1 Adjusting anomalous near-shore temperature values	27
2.4.2 Imputing and interpolating missing temperature data	28
2.4.3 Interpolating missing chl-a and turbidity values	29

2.5 Jupyter Notebook Implementation	32
2.5.1 Notebook Description	32
2.5.2 Notebook Outline	32
2.5.3 Key Outputs	34
2.6 User Notes	34
2.6.1 Summary Statistics	36
2.6.2 Summary plots	37
 Chapter 3: Correlation of chl-a with temperature and turbidity on Utah Lake	 44
3.1 Introduction	44
3.2 Analyses	47
3.2.1 Generalized least squares regression model	47
3.2.2 Outlier analysis of chl-a on turbidity levels	49
3.2.3 Temporal temperature analysis	51
3.3 Results	52
3.3.1 Loose correlation of chl-a with turbidity varies temporally and spatially	52
3.3.2 Chl-a outliers associated with low turbidity values	55
3.3.3 First algal bloom often preceded by a large temperature increase	57
3.4 Discussion	61
3.4.1 Overall correlation of chl-a with turbidity	61
3.4.2 Correlation of outlier chl-a values with turbidity	64
3.4.3 Correlation of chl-a with temperature	65
3.5 Conclusions	66
 Chapter 4: Spatial and Temporal Algal Bloom Dynamics on Utah Lake	 68
4.1 Introduction	68
4.2 Composite image generation and analysis	68
4.2.1 Image processing in GEE	70
4.2.1 Visual classification data	71
4.2.2 Composite Images	73
4.3 Composite imagery results	76
4.3.1 Blooms and chl-a plumes more frequent in shallow bays	76
4.3.2 Composite Images	77

4.4.3 Spatial differences in chl-a value distributions	82
4.5 Discussion	83
4.5.1 Spatial variation in bloom frequency and intensity	83
4.5.2 Temporal patterns	84
4.5.3 Extreme value analysis	86
4.6 Conclusions	87
Chapter 5: Conclusion	89
5.1 Key Findings	90
5.2 Error and limitations	92
5.3 Recommendations	95
References	97
Appendix A: Chl-a and turbidity plots by month	106

Chapter 1: Introduction

This chapter describes the motivation for this study, provides background on Utah Lake, introduces the study methods, and contextualizes this research within the field of water quality remote sensing and the existing body of research on Utah Lake.

1.1 Utah Lake and Algal Blooms

Harmful algal blooms (HABs) are a growing global concern due to their detrimental impacts on human health and aquatic ecosystems. These blooms, which involve excessive growth of green algae and sometimes toxin-producing cyanobacteria, can produce hypoxic and/or toxic conditions that disrupt the aquatic food web, pose health risks to humans and animals, and degrade the waterbody's economic value [1,2]. Most prevalent during warm weather, HABs are exacerbated by global climate change that raises summer temperatures and lengthens the growing season [3,4]. There is currently great interest in understanding drivers, trends, and patterns in algal growth in surface waters so that effective mitigation strategies can be developed, particularly for lakes with significant ecological and economic importance.

Utah Lake (highlighted in blue in Figure 1.1) is just such an ecologically and economically vital lake—a large, shallow, naturally eutrophic freshwater lake in Utah, USA, and a central feature of and critical water resource for the semi-arid Utah Valley. Utah Lake provides many ecosystem services to the surrounding environment and human population, including water storage, recreation, wildlife habitat, and climate regulation [5,6].

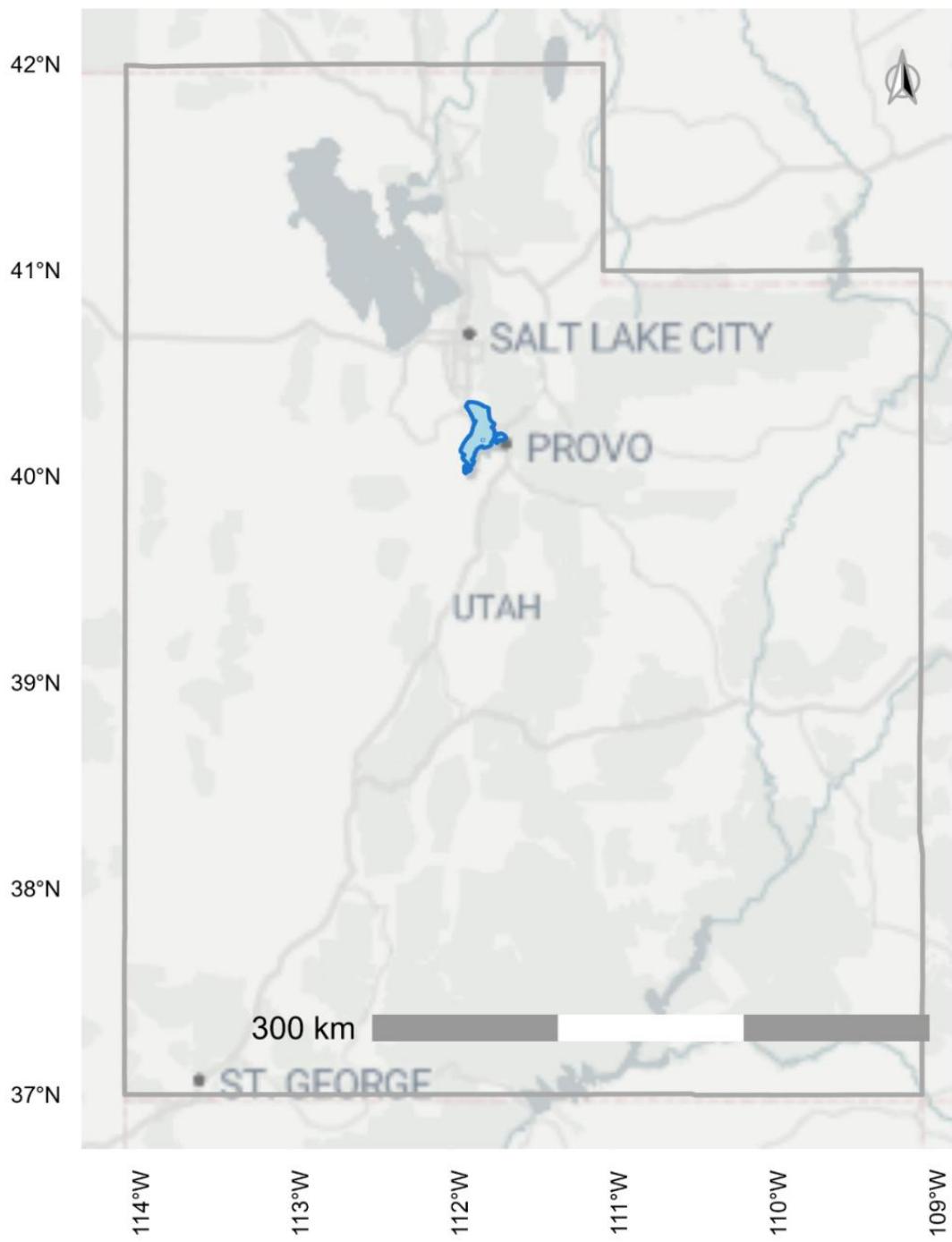


Figure 1.1: Map of the state of Utah, USA, with Utah Lake highlighted in blue.

The lake is large, approximately 40 km (24 mi) by 21 km (13 mi), and shallow, with an average depth of 3 m (9 ft). Its waters are warm, turbid, eutrophic, slightly saline, and well-oxygenated with

high pH levels—characteristics that quickly degrade and stabilize pollutants [7]. The high turbidity is mainly driven by resuspension of soft benthic sediment by intense wave action, as well as carbonate precipitation [8]. The lake's massive surface area relative to its volume causes high evaporative losses, estimated to be as much as 50% of its inflow, which results in extremely high levels of dissolved solids (>1000 mg/L) and high alkalinity[7,8].

The lake's unique geochemistry and ecological importance have led to a considerable body of work, from published academic articles to in-depth reports for governmental agencies and private-sector projects, devoted to understanding its characteristics and function [6,7,9-36].

Geologic and geochemical studies have found extremely high concentrations of phosphorus, a key nutrient, in the water column and benthic sediments of Utah Lake, likely driven by large natural sources such as stream inflows and atmospheric deposition in addition to the smaller contributions from wastewater treatment plants. These studies suggest that the lake's physical properties make the lake a "sink" for phosphorus, with up to 60% of the total load remaining in the lake, raising concentrations well above growth-limiting levels [9,10,12,15-17,35].

Studies of algal blooms on Utah Lake have found numerous unique species and observed that algal concentrations are highly variable and influenced by a multitude of environmental conditions [25,26,29,31]. One study found that algal concentrations inside of in situ microcosms responded to nutrient additions, although this effect was not universal and varied spatially [37]. In addition, there is a close link between water temperature and concentrations of cyanotoxins produced by some HABs, suggesting that warmer temperatures facilitate toxin production by these blooms [37]. Studies have also shown extremely high spatial and temporal variability of algal blooms on seasonal, intra-seasonal, and even daily time scales [25,27]. One earlier in-situ study on Utah Lake and a few studies on other large, shallow, turbid lakes have lent credence to the hypothesis that, given Utah Lake's distinct lack of nutrient limitation, water temperature and light availability as

mediated by turbidity may be more influential drivers of algal growth [30,38-41].

Other work has focused on the abundant and diverse wildlife supported by the lake, including endemic species like the threatened June Sucker and invasive species like the common carp, which was introduced in the 1800s [18,23,32].

In addition, the following factors have prompted significant research on anthropogenic influence on the lake: the lake's location at the heart of the rapidly-urbanizing Utah Valley, the recent explosion of the surrounding population and accelerating land development, and a history of industrial waste and raw sewage inflows as well as problems with invasive species [5,6,8,24,35,42].

While Utah Lake's warm, nutrient-rich waters have likely always been a highly productive environment [19], public and environmental health concerns related to these blooms have spiked in recent years. Since 2016, Utah Lake's public beaches have been closed for weeks at a time each summer due to the detection of HABs [34], raising concerns over human health and the ecological health of the lake and exacerbating negative public perception of the lake. HABs on Utah Lake deter recreation, jeopardize the aquatic ecosystem and, in the case of cyanobacteria-containing blooms, threaten human and animal health and the many acres of farmland that are irrigated with Utah Lake water [43].

The Utah State Legislature has demonstrated that addressing algal blooms in Utah Lake and other Utah waterbodies is an urgent priority—in fiscal year 2022, \$1.4 million was allocated to research, monitoring, and mitigation of algal blooms, compared to less than \$400,000 allocated in the previous fiscal year (2021) [44]. In addition, implementation of new rules targeting nutrient loading from wastewater treatment plants (WWTPs) may cost up to \$1 billion [45], though their efficacy is still a matter of debate because of nutrient loading from other sources including agricultural and stormwater runoff, atmospheric deposition, and in-lake sediment resuspension [46]. An accurate understanding of the drivers and dynamics of the algal blooms is essential to predicting the success of various proposed strategies and maximizing the benefit to the lake from available resources.

To better inform such decisions, we investigate the relationships between algal blooms and two environmental factors known to strongly influence algal growth: temperature and light availability. We also analyze temporal and spatial dynamics of algal blooms on the lake. We did not include nutrient data in our analysis other than some generalized estimations about spatial differences in the lake because prevailing nutrient levels in Utah Lake are so high throughout the growing season that they are likely not the only or even main limiting factor on algal growth [8,19,47]. For this reason, we assume that the large-scale trends and patterns in algal growth which we analyze are mainly affected by temperature and light availability.

Prior studies investigating algal bloom dynamics and the effects of turbidity and temperature on algal growth have mainly relied on mesocosm experiments and water sampling data [38-40,48-50]. These types of studies are valuable for modeling detailed interactions and characterizing specific processes, but they provide limited insight into algal growth at large spatial and temporal scales [51], and few studies have examined algal growth at a system-level scale on Utah Lake.

1.2 Remote sensing of algal blooms

The use of remote sensing data for water quality studies began in the 1970s with the first Landsat mission [33], and is an increasingly popular and thoroughly validated method for monitoring and researching algal blooms in inland waterbodies [52-54]. Data from earth observation satellites are more useful than in situ data for examining large-scale trends, patterns, and interactions, because imagery from earth observation satellites typically has greater spatial and temporal extent [55,56]. Water quality remote sensing studies provide a valuable complement to the more granular insights from in situ studies.

Remote sensing satellites allow us to estimate water quality parameters like turbidity and algal biomass by measuring the intensity of reflected light in specific wavelengths corresponding with the spectral characteristics of water, suspended sediment, and

chlorophyll-a (chl-a)—a plant pigment commonly used as an index for algal biomass [57]. Satellite imagery can provide these measurements for an entire lake over multiple years or even decades, enabling a system-level analysis of potential interactions during different seasons and under different environmental conditions [58]. Unlike physical water samples that are limited to point measurements from single locations and historical data which tend to be temporally sparse, satellite imagery provides spatially and temporally comprehensive data at the scale necessary to investigate ecological drivers and relationships between water quality parameters for an entire lake [51]. This is a key component for understanding the behavior of algal blooms in a waterbody.

The extensive coverage of remote sensing data is particularly useful for Utah Lake because its pancake-like, nearly endorheic morphology results in an extremely slow flushing rate and limited lateral mixing [36] which cause significant spatial heterogeneity in ecological and hydrologic characteristics across the lake. For example, the hydrologically isolated and heavily vegetated Provo Bay often has much clearer water than the deeper, windswept center of the lake [13]. In situ water samples, limited to single points, have difficulty characterizing such spatial heterogeneity, as well as the context and time-history provided by remote sensing data.

Previous remote sensing studies on Utah Lake have yielded valuable insights into the long-term behavior of algal blooms on the lake using images from NASA's long-running Landsat mission, including one of the very first published Landsat papers [33]. Two more recent Landsat studies both found slight decreases in overall bloom intensity and frequency since the 1980s [13,59], although Provo Bay, a shallow bay on the east side of the lake, shows worsening bloom conditions, and the drivers behind these trends remain poorly understood. In addition, these studies are limited by Landsat's revisit time of 8 days (16 days before 2013), which does not capture changes in algal blooms that often occur on a daily time scale. More information is needed to understand the trends identified by previous studies and to anticipate how a changing environment and climate may affect blooms.

The European Space Agency's Sentinel missions now provide imagery at a higher spatial and temporal resolution than Landsat (10-meter pixels vs 30-m pixels and a 2-3 day revisit time vs 8-16 day). Usable Sentinel imagery of Utah Lake begins in 2019, which means there is now a six-year dataset of relatively high-frequency imagery on Utah Lake that can be used to investigate intra-seasonal dynamics and algal bloom patterns on a shorter time scale. The high-frequency imagery available from Sentinel 2 provides a unique and unexplored opportunity for analyzing large-scale algal bloom patterns and drivers on Utah Lake.

Although the Google Earth Engine platform has greatly reduced the level of expertise and computational resources necessary for conducting water quality remote sensing studies [60], transforming large, complex satellite images into usable water quality data is still a work-intensive process. Working off of several years of prior algal bloom remote sensing research [11,13,14,60-63], we developed a method for creating an accurate but highly accessible dataset of water quality measurements from the ESA's Sentinel-2 and NASA's MODIS satellite imagery [28] using GEE. We describe this process and implementation and make the data available for Utah Lake, but it can be readily adapted to most other freshwater lakes.

Leveraging this comprehensive six-year dataset and additional data derived from Sentinel 2, this work focuses on long-term, large-scale spatiotemporal patterns, short-term trends, and seasonality in the intensity and location of algal blooms on Utah Lake and the relationship between algal growth and water column turbidity and temperature. We characterize when and where blooms typically occur, their growth and decline rates, their typical durations and return periods, and their correlation with temperature and turbidity levels. We explore the interactions and relationships between these parameters as potential drivers of algal blooms on Utah Lake. We use the patterns and trends identified by this analysis to provide insights into the potential efficacy of proposed bloom mitigation and water quality management strategies for Utah Lake. In addition, we demonstrate an accessible method for generating accurate water quality datasets from earth observation satellite imagery which can be used to study eutrophication and

HABs on other freshwater lakes. We make the Utah Lake data available for other researchers and water managers concerned with Utah Lake to perform their own analyses.

Chapter 2: A six-year, spatiotemporally comprehensive Utah Lake chl-a dataset

This chapter is based on the article “A Six-Year, Spatiotemporally Comprehensive Dataset and Data Retrieval Tool for Analyzing Chlorophyll-a, Turbidity, and Temperature in Utah Lake Using Sentinel and MODIS Imagery,” published in *MDPI Data* [64].

2.1 Introduction

While Google Earth Engine has greatly simplified the work required to conduct remote sensing analyses and reduced the associated computational needs, generating usable data from earth observation satellites still requires significant technical expertise and time-intensive processing steps. We developed this method of downloading and processing imagery from Sentinel 2 and MODIS into a convenient, accessible csv format and provide the resulting dataset along with the method implemented in a Jupyter Notebook to make remotely-sensed water quality data more accessible to researchers, water managers, and others interested in Utah Lake and to facilitate the use of satellite data for those interested in applying remote sensing techniques to other waterbodies.

We extended methods from Cardall, Tanner and Williams [11] to generate the data using the GEE Python API. While tools and methods exist for processing all pixels contained in a lake or waterbody, these datasets are massive—millions of pixels per image and large numbers of images per year. For most analyses, a subset of pixels, randomly selected to represent lake processes, can be used to characterize lake processes and spatial patterns. We generated three different sets of 200 random points: one that covers the whole lake, one with 50 points each selected from four boxes

placed in areas of the lake where we expected water conditions to vary, and one with 100 points in “near-shore” areas of the lake and 100 points in “open-water” areas. At each point we extracted the associated band values, which represent the measured intensity at different wavelengths of light in each pixel, from every usable image in the collection. The point subsampling method allowed us to obtain data even from partially clouded images, which is one of the major difficulties associated with remote sensing studies, because we extracted values from individual pixels only if they were unaffected by clouds or cloud shadow.

We used methods adapted from Cardall, Hales, Tanner, Williams and Markert [14] to create empirical models fitted using in situ data and near-coincident satellite imagery to estimate chl-a and turbidity values. We acquired daily daytime and nighttime water temperature measurements at each point from NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS) imagery. We performed quality assurance checks on the data and interpolated the data to a daily timescale for use in statistical time series analyses.

We created this comprehensive water quality dataset for Utah Lake for use by interested researchers and water quality managers. We provide both the dataset and tools used to generate it as an example of how to generate and use this type of data from satellite imagery and to allow others to easily adapt these tools to different locations. These data and the summary statistics represent highly accessible and comprehensive information about Utah Lake water quality and demonstrate an efficient method to access these types of remotely sensed image products and to generate reasonably sized data sets for analysis that capture both spatial and temporal patterns.

2.2 Data Description

The three datasets are comprised of daily values for chl-a ($\mu\text{g/L}$), turbidity (NTU), and daytime and nighttime water temperature ($^{\circ}\text{C}$) either derived from band values from a satellite image or interpolated from surrounding data. The three sets of sampling data are formatted as a single csv file with the following columns:

- **date**: daily timestep from January 1st, 2019 to March 20th, 2025, formatted as mm/dd/yyyy.
- **point_id**: a unique integer identifier for each sample point location
- **latitude**: in the WGS 84 projection
- **longitude**: in the WGS 84 projection
- **dataset**: a label indicating the set of sampling points that datapoint belongs to; whole-lake, boxes, or clusters (see section 2.4).
- **category**: for the clusters and boxes datasets, a label that indicates which sub-category the datapoint belongs to. For clusters, either Open Water or Near Shore; for boxes, one of Provo Bay, North Lake, Center Lake, or Goshen Bay.
- **in_PB**: FALSE if datapoint is outside Provo Bay, TRUE if inside.
- **int_flag**: FALSE if datapoint is from a satellite image, TRUE if datapoint is interpolated from surrounding data (see section 2.4). Note that this flag only applies to chl-a and turbidity data, not MODIS-derived data.
- **parameter**: chl-a, turbidity, dayTemp, or nightTemp. chl-a and turbidity are derived from Sentinel 2 imagery; dayTemp and nightTemp are derived from MODIS imagery.
- **value**: computed parameter value for the image pixel in the location specified by the coordinates on the specified date, or the interpolated value if satellite data for that pixel was not available on that date. The units for chl-a, turbidity, and temperature are $\mu\text{g/L}$, NTU, and $^{\circ}\text{C}$, respectively.

There are 5,611,432 rows in total, with 468,400 values for each of the three datasets and four parameters.

We published a second csv file that includes the red, green, blue and red edge band values extracted from GEE, which we used to compute the chl-a and turbidity values. We chose to publish this as separate data to make the main dataset as user-friendly as possible, but the values can be easily joined with the main dataset using the date and point_id fields.

- **date**: daily timestep from January 1st, 2019 to March 20th, 2025, formatted as mm/dd/yyyy.
- **point_id**: a unique integer identifier for each sample point location
- **red**: value of the red band
- **green**: value of the green band
- **blue**: value of the blue band
- **RE1**: value of the red edge 1 band

2.3 Methods

2.3.1 *Sentinel image processing*

We extended methods developed by Cardall, Tanner and Williams [11] to download and process imagery from the Sentinel 2 mission using GEE [65]. We used the harmonized Sentinel 2 MSI Level-2A orthorectified and atmospherically-corrected surface reflectance image collection available on the Google Earth Engine catalog. The Sentinel 2A and Sentinel 2B satellites provide images every 2-6 days starting in 2015, but the first usable images of Utah Lake are from 2019. This results in a collection of 937 usable images of Utah Lake collected roughly every 2-3 days (with larger gaps occurring mainly during the winter when cloud cover is more frequent) from 2019-2024. We accessed these images using tools built on the GEE Python API using Python code implemented in a Jupyter Notebook.

We used the Sentinel Scene Classification Layer (SCL) to filter pixels contaminated with clouds, cloud shadow, sensor failures, or

other issues, and pixels that contained land, snow, or ice. We kept all pixels with an SCL value of 4, 5, 6, or 7, which represent vegetation, bare soils, water, and low probability clouds, respectively. We chose to include the soils and vegetation categories because visual inspection of images showed that extremely turbid areas of Utah Lake were sometimes misclassified as soils in the SCL layer, and very intense algal blooms were sometimes classified as vegetation. Because we obtained data from the satellite imagery by extracting pixel values at known sample points, and our sample points were all in locations on Utah Lake that are known to be covered by water (see Section 2.4 for details), this method allowed us to preserve the most data possible and not erroneously exclude valid data from extremely turbid water or intense algal blooms. In addition, we masked each image using a 50% occurrence threshold from the JRC Global Surface Water Mapping Layers dataset [66] to further ensure that only pixels we could reasonably expect to be land-free were included in our dataset.

2.3.2 Remote sensing models

Remote sensing models provide information about water column conditions by measuring light reflection from optically-active water column constituents such as chl-a (a plant pigment commonly used as an index for algal biomass) and suspended solids. Each of these has a unique “spectral signature”—absorbance and reflectance peaks at specific wavelengths. Multispectral satellite imagery measures the intensity of earth-leaving radiance in a set of spectral bands selected specifically to characterize various materials, such as vegetation and soils (which can be used for turbidity). We use the intensity of the radiance in the bands that correspond with a material’s spectral signature to generate models which estimate the amount or concentration of that material in a pixel [67].

For some applications, physics-based models built on known spectral characteristics of a material provide sufficiently accurate results; however, for optically complex waters these models often fail due to the presence of other materials that have overlapping spectral signatures or cause unpredictable scattering of reflectance

[13]. Turbid, productive lakes, like Utah Lake, contain optically active constituents with absorption features that overlap those of chl-a. For example, colored dissolved organic matter (CDOM) and detritus reflect strongly in the same blue-green area of the electromagnetic spectrum as chl-a, so empirical models that rely solely on blue and green spectral channels cannot provide accurate estimates of chl-a when CDOM and detritus are present [68]. The spectral features of algal blooms also vary with the species dominating the bloom and other water column chemistry conditions, so models based on in situ data from the study location provide more accurate results than a generalized physics-based models. [69].

In these cases, empirical regression models based on data pairs of in situ measurements and coincident or near-coincident satellite imagery provide better results. These models are fit using data from the image bands. The regression analysis quantifies relationships not apparent from spectral characteristics alone and allows us to capture relationships that are not yet understood [14]. This means that a model based on empirical data is better able to identify unique, specific spectral characteristics of a material in a certain area.

Empirical models based on observed data are well-suited to Utah Lake for two reasons: first, Utah Lake is extremely optically complex due to high turbidity levels and a high diversity of algal species, which vary throughout the season and year to year [70], making it difficult for a physics-based model to adequately characterize chl-a and turbidity. Second, because of its environmental and economic importance, there is a long-running and relatively comprehensive history of water samples collected on Utah Lake, providing sufficient data to generate accurate empirical models.

2.3.3 *In situ* data

To build our models, we acquired in situ Utah Lake water quality measurements from the Utah Department of Environmental Quality's Ambient Water Quality Monitoring System (AWQMS). We obtained 752 measurements of turbidity and 1,937

measurements of chl-a taken at various locations on the lake from 1978 to 2022.

We then used GEE to identify and process Sentinel images of Utah Lake taken within 12 hours of an in situ sample collected on Utah Lake. We identified the pixel associated with the in situ sample location and extracted the image band values to generate “matchups”—sets of in situ measurements and near-coincident satellite image data. We found 154 pixel matchups for chl-a and 113 pixel matchups for turbidity. The locations and number of matchups at each location are shown in Figure 2.1. The lake is fairly well-represented in the data, although there are more data points close to the shoreline than in the open lake.

2.3.4 Combining physics-based and empirical models

Using the matchups dataset, we evaluated various methods for developing empirical models of chl-a and turbidity for Utah Lake. One method used machine learning to generate models using a near-exhaustive set of bands and band combination values as potential model features [14]. Another method used in situ data to improve a physics-based normalized difference index model (the difference of a pair of bands divided by the sum of those bands) by fitting a site-specific coefficient to the index value [68]. Ultimately, a combination of these methods produced the most accurate models for both chl-a and turbidity on Utah Lake.

We tested regression models created using Least Absolute Shrinkage and Selection Operator (LASSO) regularization, which selects model terms from a large feature space of all possible band combinations and modifications (such as the product of two bands, or the inverse or square of a band) [14]. The LASSO models were prone to overfit and included an excessive number of terms with very large or very small coefficients and non-random errors. They were helpful, however, in identifying potential model features.

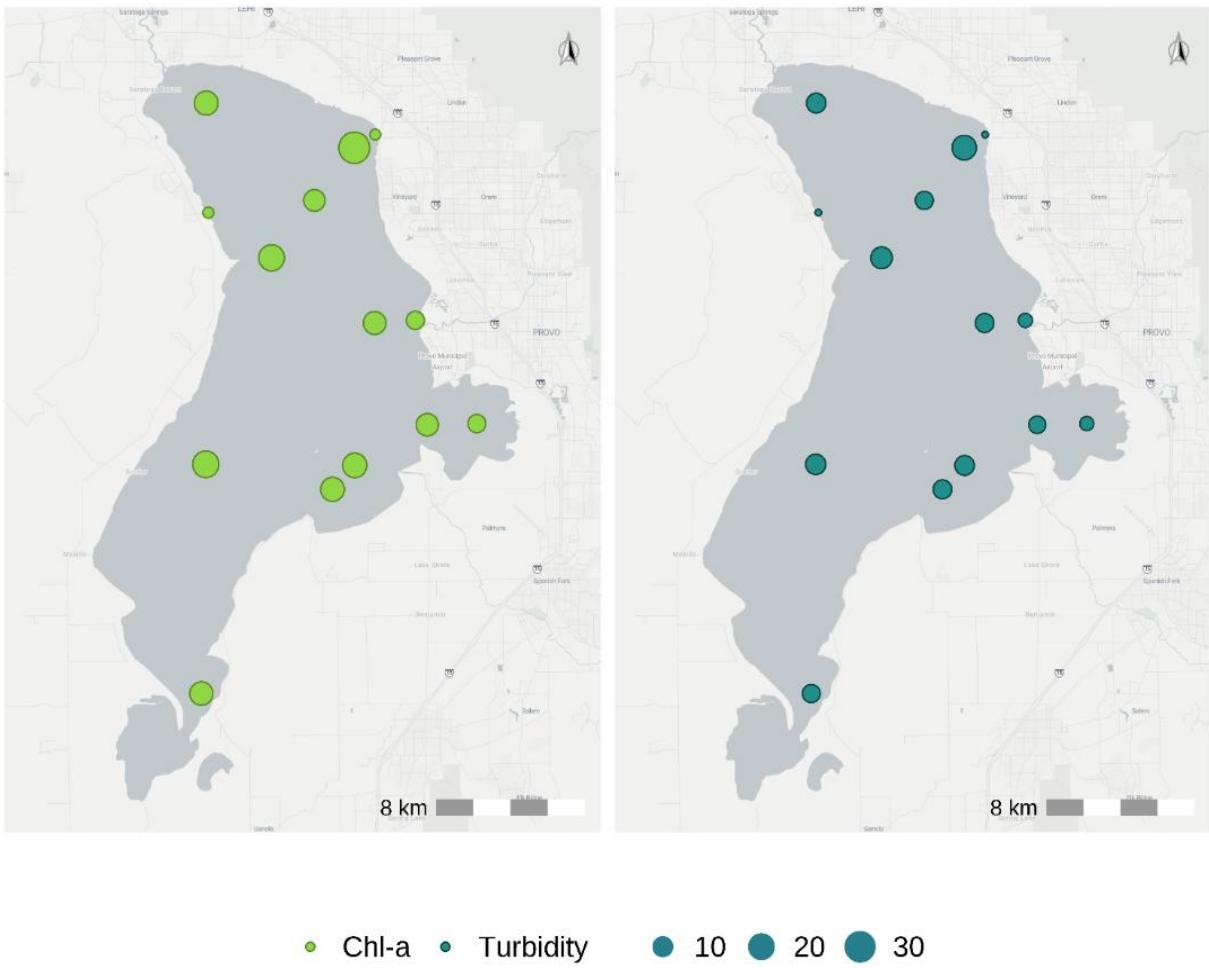


Figure 2.1: Map of locations represented by pairs of near-coincident Sentinel-2 satellite and in situ water quality measurements between 2019 and 2022. In situ data were collected by the Utah Department of Environmental Quality. Dot size indicates the number of datapoints at a location for which there was a near-coincident Sentinel 2 image.

The normalized index models we tested generated good estimates when fitted with coefficients based on the in situ data, but introduced a limit on predicted values at the function maximum, which was significantly lower than the highest in situ data values. To address this, we added band values identified by the LASSO models as being well-correlated with the target parameter as additional model features or terms, which increased the accuracy of the models and allowed them to predict values over the full expected data range.

We fit the models using the `LinearRegression` function from the scikit-learn Python library [71]. We evaluated the accuracy of the

models using the coefficient of determination (R^2) and root mean square error (RMSE) values. For aquatic remote sensing models, R^2 values between 0.6 and 0.8 are commonly accepted [72], while evaluation of RMSE values depends on the range of the value being modeled. Retana *et. al.* [72] achieved an R^2 of 0.98 with an empirical model for TSS

For chl-a, the range of measured in situ values was 254 $\mu\text{g/L}$, with a standard deviation of 35 $\mu\text{g/L}$, and mean of 24 $\mu\text{g/L}$. Since the error of the in situ chl-a measurements was likely ~10% or greater [73], we considered models with RMSEs below 2.5 (i.e., within the average error) to be useful. For turbidity, the range of measured values was 158 NTU, with a standard deviation of 36 NTU, and a mean of 25 NTU. The error for the turbidity measurements is much lower than for chl-a—around 0.3 NTU—but we considered models with RMSEs below 5 NTU, which is a small value relative to the range and average of the data, to be useful.

2.3.5 Chlorophyll-a

Although season-specific chl-a models have been shown to be more accurate on some lakes [74], the limited amount of data available and the extreme inter-annual variability of Utah Lake as a basin-bottom lake and a managed reservoir made this approach impractical. Season-specific models also require a significant amount of in situ data to characterize behavior in the different seasons. We were unable to generate useful models if we subsetted the data into seasonal groups because of the limited amount of data for each season, so we used a whole-season chl-a model based on data collected between April and September.

We used methods similar to those described by Cardall, *et. al.* [14] to analyze a number of potential model features and develop a sparse, interpretable chl-a model. We found that models based on the Normalized Difference Chlorophyll Index (NDCI) developed by Mishra and Mishra [68] performed better than models based on other indices or on plain band values. The NDCI is calculated as:

$$NDCI = \frac{\rho_{red} - \rho_{RE}}{\rho_{red} + \rho_{RE}}, \quad 2.1$$

where ρ_{red} and ρ_{RE} represent the red and red-edge band values, respectively.

We fit the linear regression coefficients including terms for the NDCI value and the square of the NDCI value. This model provided acceptable results (R-squared values around 0.8), however it introduced an artificial limit on predicted chl-a values at the maximum of the equation. The maximum from this initial equation was approximately 110, which is much lower than the highest chl-a values we expected based on in situ measurements, some of which were as high as 300 µg/L. To address this, we added two band value terms to the linear regression equation; this allowed the model to predict chl-a values over the full expected range.

The final adjusted NDCI model was:

$$\ln(chla) = 2.90 - 8.95(NDCI) - 17.20(NDCI^2) \\ - 21.96(\rho_{blue}) + 17.11(\rho_{green}), \quad 2.2$$

where ρ_{blue} and ρ_{green} are the red and green band intensity values, respectively. This model produced an R² value of 0.80 and an RMSE of 0.48 (an RMSE of 1.62 when unlogged).

3.3.4 Turbidity

We used a similar modeling approach—a combination of a normalized difference and unmodified band values—to produce the turbidity model. This hybrid model produced better results for turbidity than other types of models. We tested the normalized difference of every pair of bands available from Sentinel 2 and found the blue-red pair had the best correlation with Utah Lake turbidity values. We called this the Utah Lake Normalized Difference Turbidity Index (ULNDTI), and calculated it as:

$$ULNDTI = \frac{\rho_{blue} - \rho_{red}}{\rho_{blue} + \rho_{red}}, \quad 2.3$$

where ρ_{blue} represents the blue values.

A turbidity model with ULNDTI and squared ULNDTI values as the only terms created an artificial limit on the predicted values (the same issue we found with the chl-a model), so we tested various plain band additions to the model to eliminate this limit. We found that adding a single red-edge band value to the model produced the best results, and the final ULNDTI model was:

$$\ln(turbidity) = 2.47 - 2.73(ULNDTI) + 0.21(ULNDTI^2) + 10.70(\rho_{RE}) \quad 2.4$$

which produced an R-squared value of 0.89 and an RMSE of 1.25 (3.50 unlogged).

We generated data based on this model excluding model estimates of turbidity with values above 500 NTU because visual inspection of the satellite images indicated that these high values were associated with locations where water levels were much lower than normal. This meant that the pixel contained both water and land or that the water was shallow enough to image the lake bottom through the water column—both of which would invalidate the data point. Based on the in situ data, we would not expect to see turbidity values above 500 NTU.

3.3.5 Temperature

We provide temperature values from the Lake Surface Water Temperature (LSWT) product, derived from the MODIS Global Daily Surface Temperature and Emissivity 1km image collection from the Aqua satellite available on the GEE data catalog, without modification. Numerous studies have validated the use of LWST data where in situ lake temperature data are not available [75-80]. Studies of LWST data on the Great Salt Lake, which is near Utah Lake, and Lake Taihu in China, both of which share many characteristics with Utah Lake, found a cold bias in LWST estimates for those lakes [81,82], and such a bias may exist for Utah Lake.

However, Lazhu *et. al.* [75] point out that the observed biases may come from a failure to consider the representativeness of in situ samples for the entire area of these large, shallow lakes with slow lateral mixing. In addition, since the use cases for these data are concerned mainly with trends, a consistent bias in one direction (as would likely be the case on Utah Lake) should not affect the validity of these data for analyzing and characterizing trends or changes in lake conditions. Nevertheless, the results of analyses based on MODIS temperature data should be checked against in situ data whenever possible and the possibility of the cold bias should be considered.

We observed that a random portion of the MODIS-derived temperature measurements on Utah Lake were outside the reasonable range of water temperatures, which is to be expected with temperature measurements based on surface emissivity, but the overall distribution of MODIS temperature data matched the distribution of in situ temperature measurements (see Figure 2.12). We chose not to exclude or adjust the out-of-range temperature measurements since they are still useful for correlation and trend analyses.

Because MODIS pixels are large, (~1km on each side), pixels near the shoreline of the lake include both land and water, which means the temperature values in these pixels are not representative of the water temperature. To eliminate the effects of mixed pixels in our data while including temperature for the near-shore areas, we used a nearest-neighbor spatial interpolation technique to replace values from mixed pixels with values from the closest “pure” pixel. Section 2.4.1 provides details and justification for this approach.

2.4 Image sampling and model application

Remote sensing studies performed on large areas and over longer time scales use data aggregation or subsampling to extract meaningful insights from large amounts of spectral image data. The method and scale of aggregation or subsampling depends on the characteristics of the study location—if it is largely homogenous, then an aggregate value, such as the mean or median, for the entire

area may suffice, but for areas with high heterogeneity it is necessary to aggregate over smaller areas or use subsamples. We assumed that Utah Lake is spatially heterogeneous relative to most other waterbodies because it is not well-mixed laterally—there is significant variation in water column conditions throughout different areas of the lake, especially when algal blooms are present [36]. To characterize the spatial heterogeneity of the lake while also maintaining the dataset at a manageable size, we subsampled the image data by extracting the pixel values at predefined sets of points from each image rather than calculating aggregated statistics. This subsampling technique had the additional benefits of:

- Making the data suitable for statistical analyses that assume a random sample
- Allowing us to extract more usable data from images with partial cloud cover
- Eliminating the need for complex and imprecise water masking procedures, because we located sample points only in areas where we knew there would be water.
- Providing a smaller, more accessible dataset relative to the extremely large and complex dataset of satellite imagery over extended time periods.

We created three datasets for the lake with the following subsample approaches: one subsample dataset represents the entire lake area, one dataset represents four different areas of the lake we expected to exhibit different behavior based on previous research [13], and one dataset represents the lake divided into near-shore and open-water regions. We identified these near-shore and open-water regions with a machine learning clustering algorithm. The published data includes all three datasets and a flag identifying the dataset to which each measurement belongs.

2.4.1 Whole-lake samples

The shoreline of Utah Lake fluctuates significantly with water level, so we could not use a geographically defined shoreline or the shoreline from a single image as the boundary for the set of whole-

lake sample points. Instead, we created a shoreline boundary by building a composite image of all Sentinel 2 images of Utah Lake taken during the growing season (April – October) and then applying the modified normalized difference water index (MNDWI) to the composite image and defining pixels with an MNDWI value above 0.25 as water. This provided a minimal lake boundary for the time period covered by the Sentinel 2 images used in this dataset. We then generated 200 data points with locations randomly selected from within this lake area.

Figure 2.1 shows the 200 randomly generated sample points in the whole-lake collection. There are no sample points at the very south end of the lake in the shallow part of Goshen Bay. Although it is defined as part of Utah Lake on most maps, during the time period covered by Sentinel 2 data set, this area was dry, so it did not register as water in the MNDWI-thresholded composite image. Therefore, our Sentinel 2-specific lake boundary, which otherwise matches basemap boundaries quite closely, does not include the southernmost area of Goshen Bay, which is correct for our study period.

2.4.2 Boxes sample

We defined four 1.5km (0.93 mile) boxes (150 hectares) in areas of the lake where we expected different water column conditions based on prior research and knowledge of the lake (Figure 2.2): one at the north end of the lake near the Jordan River outflow (North Lake); one in the center of the lake (Center Lake); one at the south end of the lake just above Goshen Bay (Goshen Bay); and one in the center of Provo Bay on the east side of the lake (Provo Bay). Provo Bay is shallow (less than a meter deep on average), hydrologically isolated from the rest of the lake, receives a large portion of the anthropogenically-impacted inflow to the lake, is heavily vegetated, and experiences intense and frequent algal blooms. Goshen Bay is completely open to the main lake but is also extremely shallow and surrounded by agricultural land. The Goshen Bay box is just outside the bay itself in the part of the lake nearest the bay outlet (Figure 2.2). Both Provo and Goshen Bays tend to experience more frequent and intense algal blooms than the

main body of the lake [13]. The Center Lake location represents a region in the main lake with deeper water (typically about 3 meters) that experiences greater wind and wave action than the bays. The North Lake location represents a region similar to Center Lake, but slightly less deep, and the closest to both heavily-developed land and the only outlet for the lake. Because the prevailing wind is to the north, algal blooms can be driven towards this area.

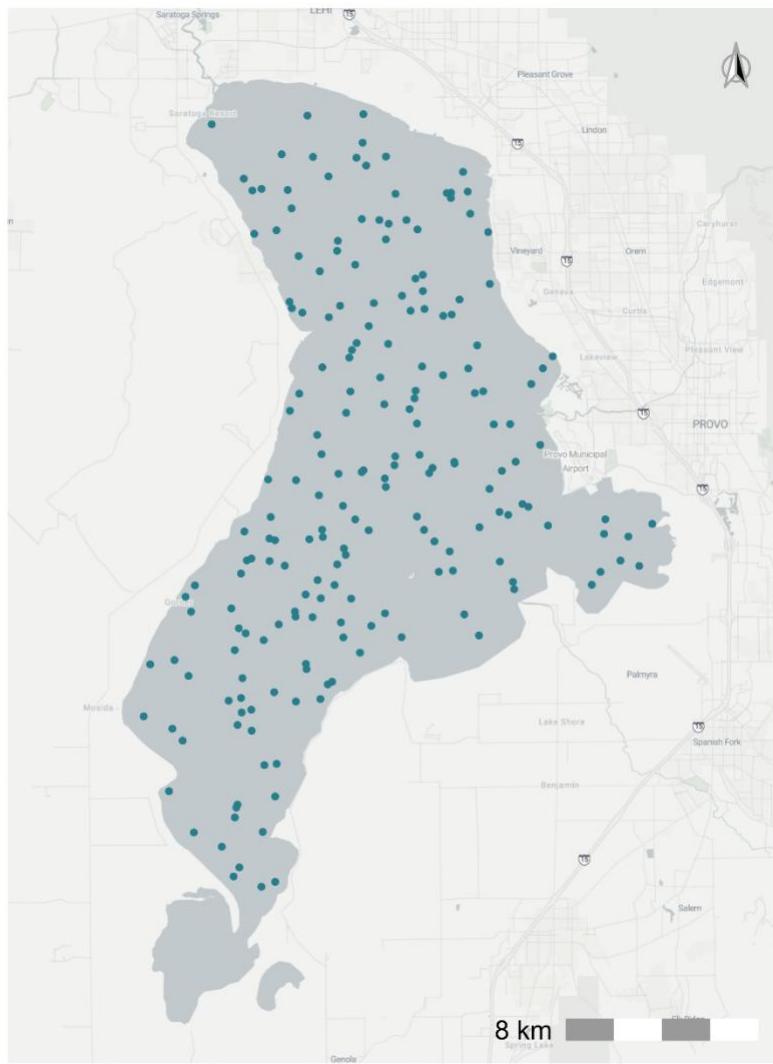


Figure 2.1: Whole-lake sampling point collection with 200 randomly-placed sample points. There are no sampling points at the southern tip of the lake in the shallow part of Goshen Bay because this area was dry during the period covered by the Sentinel 2 collection.

We generated a sampling pattern of 50 randomly located points using the GEE randomPoints function within each 1.5 km box. We used the same spatial distribution of sample points in each of the 4 boxes. The locations of the boxes and the random sampling pattern are shown in Figure 2.2.

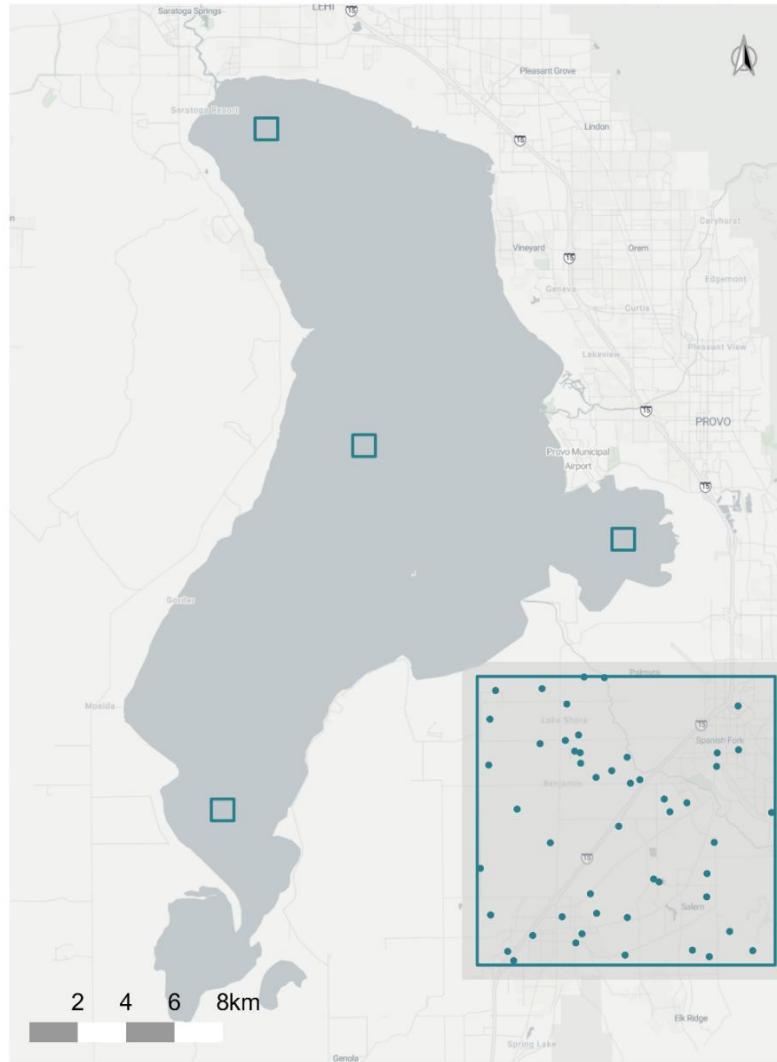


Figure 2.1: Locations of the four 1.5 km (0.93 mile) sample boxes(150-hectare area). sample boxes with inset showing the random sampling pattern used to extract 50 datapoints from each box. We used the same spatial distribution of sample points in each of the 4 boxes.

2.4.3 Machine learning near-shore, open-water clusters

Based on prior research and experience with Utah Lake, we knew that near-shore water column conditions often differ greatly from conditions on the open lake. We wanted to sample data in a way that accounted for this phenomenon, but defining what exactly is “near-shore” versus “open-water” on Utah Lake is difficult due to the lake’s shallow bathymetry and dynamic shoreline. Rather than trying to define an arbitrary boundary or buffer around the shore, we used a machine learning algorithm to identify areas where pixels tend to share characteristics over multiple images. With this clustering algorithm we delineated near-shore and open-water areas of the lake based on prevailing patterns in spectral characteristics for those areas.

We used the WekaXMeans clustering algorithm [83] in GEE because it determines the correct number of clusters itself rather than requiring the user to set the number, and we were unsure of how many distinct clusters would be present in the lake.

WekaXMeans is a K-means-based algorithm that incorporates model selection to more efficiently estimate the number and location of clusters using a Euclidean distance function. To reduce the computational requirements for the algorithm, we filtered the image collection to only include images with less than 50% cloud cover (which resulted in 598 images) and collapsed the image collection into a single multi-band image by calculating the 25th, 50th, and 75th percentile band values across the entire image collection for every pixel. The resulting image included the three values for each percentile as a band in the image for each of the original image bands—one for the 25th percentile values, one for the 50th, and one for the 75th. Thus for the Sentinel images with 10 optical bands, this results in a 30-band image. We applied WekaXMeans to these band data to identify clusters.

In multiple trials, the algorithm consistently identified three clusters within the lake (and a fourth cluster representing land), as shown in panel A of Figure . The algorithm did not assign physical meaning to these clusters, but by inspection we classified them as follows: Cluster 2 represents pixels that contained only land in

every image in the collection. Cluster 1 represents 'mixed' pixels which likely contain both land and water in a single image or contain water in some images and land in others. We chose to exclude data from Clusters 1 and 2 because estimates of chl-a and turbidity based on our models are only valid for pure water pixels. The two clusters of interest are Clusters 3 and 4, which represent open-water and near-shore pixels, respectively. Panel B of Figure 2.3 shows the random sampling pattern, with 100 points per cluster, that we used to extract data from each of these clusters. The clustering algorithm classified the majority of Provo Bay, and the area near the mouth of Goshen Bay as near-shore, likely because of the shallow nature of these bay areas, and shared water characteristics such as suspended sediments.

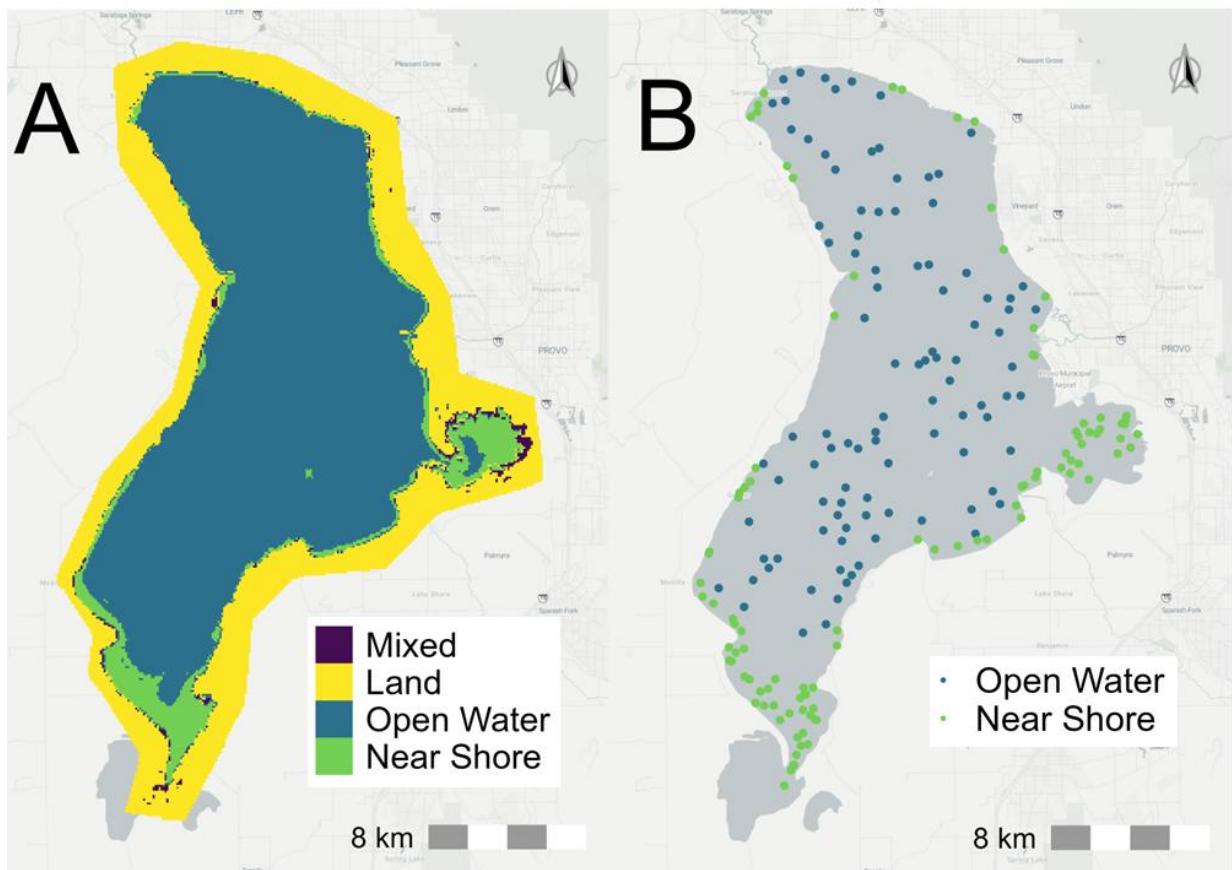


Figure 2.3: Map of sample clusters. Panel A shows the cluster areas and panel B shows the 200 sample points (100 per cluster) used to extract data in the open-water and near-shore clusters.

2.4.4 Model application

For each of the three datasets (the whole-lake sample, sampling boxes, and the clusters), we used GEE to export a csv of band values and MODIS temperature data for every sample point in each image of the collection. We then used Python to apply the models described in Section 2.3 to the band values for each data point, generating an estimate of chl-a and turbidity for each point. This was more efficient for computation and storage requirements than applying the models to the entire images in GEE and allowed us to perform some additional quality checks on our data.

2.4 Imputation of missing and anomalous values

To facilitate time-series analysis on the irregularly-spaced satellite data, we created daily data with no missing values. We used a combination of imputation with mean values (for temperature data), and Piecewise Cubic Hermitean Interpolation Polynomials (PCHIP) to interpolate temporally between data points. We used PCHIP interpolation because it honors local data limits and constructs smooth curves between data points while preserving the shape and monotonicity of the data. Unlike standard cubic splines, PCHIP avoids overshooting and produces more realistic interpolations for datasets with sharp changes or non-uniform behavior, as can be seen in the example in Figure . Figure 2.7, which compares the distributions of the original chl-a, turbidity, and temperature values from each dataset with the distributions of the interpolated values, shows that the distributions match well. This means that PCHIP generated reasonable values between observed data points that do not skew the observations higher or lower or introduce outliers. Unless otherwise specified, each of the three datasets underwent the same processing steps.

2.4.1 Adjusting anomalous near-shore temperature values

We observed that MODIS temperature values within ~1km of the Utah Lake shoreline were often unrealistically high because they came from ‘mixed’ pixels that included both land and water. We determined that this effect was entirely due to the mixed pixel

effect rather than any real pattern of warmer water near-shore by analyzing data from the AWQMS database, which showed that, although there are often spatial differences in temperature, near-shore water is not consistently warmer than water more than 1km from shore. Because there was not sufficient evidence for a consistent relationship between open-water and near-shore temperatures and we could not quantify the influence of mixed pixels on temperature data, we instead replaced MODIS temperature values collected within 1km of the Utah Lake shoreline with the value of the nearest MODIS pixel that was further than 1km from the shoreline. We identified nearest neighbors with the NearestNeighbors function from the scikit-learn package in Python [71], and checked the results after replacement by visually inspecting the data for several images and examining the distributions to ensure there were no extreme shifts as a result of the replacements.

The 1km buffer excluded the entirety of Provo Bay, but there is one location in the center of the bay with a single unmixed pixel from MODIS where it is possible to collect temperature data. We extracted temperature values for Provo Bay from that point and used these data as temperature values for Provo Bay in all three datasets. We found this to be the best way to exclude unusable data while still preserving usable information about Provo Bay, because it is hydrologically isolated and does not always move with the rest of the lake, so values taken from the open lake would not have been representative of the bay.

2.4.2 Imputing and interpolating missing temperature data

MODIS products have built-in cloud masks. MODIS pixels are much larger than Sentinel pixels, 1 km compared to 10 m, so the MODIS cloud masks usually cover a larger area than the Sentinel cloud masks, since the higher resolution of Sentinel allows more pixels to be preserved. This resulted in some datapoints with usable values derived from the Sentinel data having no associated values for temperature. MODIS does, however, provide daily data, while Sentinel data are only collected every 2 to 3 days. Since temperature values do not vary significantly across the lake (the average

difference in the minimum and maximum temperature measured on the lake over the study period is 2°C, with a maximum difference of 15°C that occurred on two days), we used the median temperature value for the lake to impute missing data rather than rely on temporal interpolation. So for any MODIS image that had at least one usable temperature measurement, but with locations missing measurements due to clouds or cloud shadow, we imputed the values for those locations with the median of the locations with usable data for that image. When no data were available (i.e., in the case of total cloud cover), we interpolated the temperature data through time using PCHIP to generate the missing values.

Table 2.1 shows the number of imputed and interpolated values for the MODIS data for the three sample data sets. Each dataset has 200 points for each time step for a total of 434,800 data points. This reflects the 2,023 days from January 1st, 2019, when usable Sentinel data starts, to July 15th, 2024.

Table 2.1: Number of values imputed and interpolated for MODIS temperature datasets

Dataset	Parameter	Values Imputed with Mean	Values Temporally Interpolated with PCHIP
Whole-lake	Day Temp	102,895 (22%)	623 (0.14%)
Whole-lake	Night Temp	97,590 (21%)	640 (0.15%)
Boxes	Day Temp	102,485 (22%)	722 (0.04%)
Boxes	Night Temp	102,522 (22%)	711(0.04%)
Clusters	Day Temp	104,462 (23%)	614 (0.04%)
Clusters	Night Temp	100,103 (22%)	426 (0.02%)

2.4.3 Interpolating missing chl-a and turbidity values

The Sentinel revisit time alternates between 2 and 3 days. Occasionally, due to partial or complete cloud cover or the presence of ice on the lake, the time gap was longer, but the vast majority of gaps in the data were either 2 or 3 days, as shown in Figure 2.4,

which we created using the set of 200 whole-lake sampling points. The few instances of longer gaps between data points nearly always occurred during the winter months, when cloud cover and ice were more frequent. We used PCHIP to temporally interpolate between Sentinel data to provide data on a daily timescale.

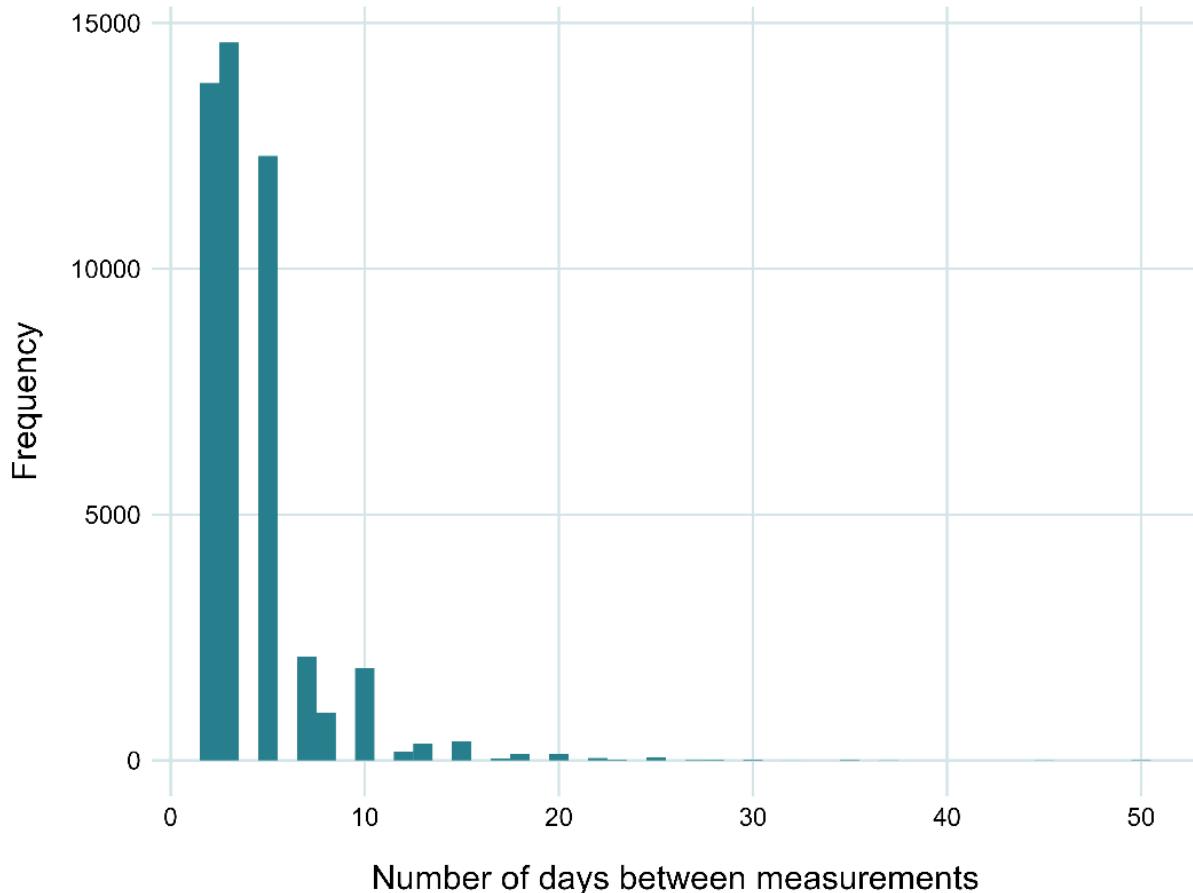


Figure 2.4: Histogram of data gap frequency (total count) for the sample points in the whole-lake dataset. These data include winter months when cloud cover and ice are more frequent.

In addition to filling in data between satellite revisit times, we also used PCHIP to fill in data from partially-clouded images. Since turbidity and chl-a can vary significantly throughout the lake on a given day, we found it was not accurate to spatially impute chl-a and turbidity values missing due to cloud cover with that day's mean value for un-clouded pixels as we did with MODIS data. Instead, we imputed missing chl-a and turbidity values using

temporal interpolation based on values of that same pixel from previous and future images.

We interpolated daily values for chl-a, temperature, and turbidity for each sample point using the PchipInterpolator function from the SciPy package in Python [84]. Figure

2.5 shows PCHIP interpolation results and demonstrates how it follows the existing data without overshooting the local points. We did not use the interpolated data for computing statistical properties such as means and medians, because PCHIP assumes a smooth transition between points, when in reality there may have been a dip or spike in actual values. We intend the interpolated data to be used mainly for time series and trend analysis where a consistent time-step is required.

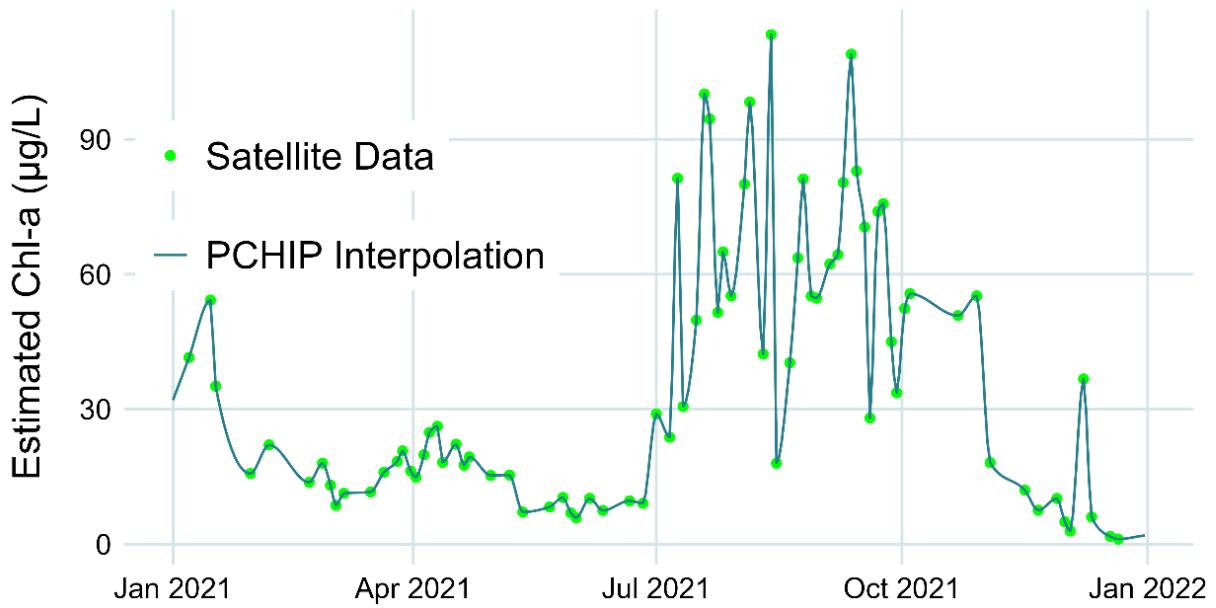


Figure 2.5: Example of PCHIP interpolation for 2021 data for the sample point at coordinates 40.14755 -111.8626. This figure contains 79 measurements and 286 interpolated points.

For the whole-lake, boxes, and clusters datasets we interpolated 376,805 data points from 91,595 observations, 385,892 data points from 82,508 observations, and 378,080 data points from 87,978 observations, respectively. There are more interpolated points than

measurements because we generated daily data from data collected only every 2 or 3 days, with occasional larger gaps due to clouds, ice, and other issues. Because of these limitations, we recommend the interpolated data be used primarily for time series analysis, while actual measurements be used for other statistical characterizations.

2.5 Jupyter Notebook implementation

2.5.1 Notebook description

We have included a Jupyter Notebook with the published paper [64] which demonstrates the workflow we used to generate the datasets previously described. This Notebook can be easily modified for other waterbodies, or to change sampling locations on Utah Lake. The following is an overview of the code which we used to implement the methods described above.

2.5.2 Notebook outline

- 1) Setup
 - a) Load packages and set up GEE API
 - b) Define rough lake outline with hand-selected coordinates (to analyze a different water body, user can supply different coordinates)
- 2) Retrieve and process satellite image collection
 - a) Load Sentinel 2 data
 - b) Apply processing functions that scale bands, perform initial quality assurance, and rename bands for future processing
- 3) Create sample point collections
 - a) Whole-lake collection
 - i) Generate a whole-lake boundary by creating a composite Sentinel 2 image and applying the Modified Normalized Difference Water Index
 - ii) Export boundary as GEE asset
 - iii) Generate sample points inside lake boundary, add metadata features, and export feature collection as GEE asset
 - b) Clusters collection

- i) Generate cluster boundaries by creating a composite Sentinel 2 image and applying clustering algorithm to computed band percentiles
 - ii) Export clusters as polygons to GEE asset
 - iii) Generate sample points inside cluster boundaries, add metadata features, and export feature collection as GEE asset
 - c) Boxes collection
 - i) Generate box boundaries with user-selected coordinates and export as GEE asset (to analyze a different water body, user can supply different coordinates)
 - ii) Generate sample points inside box boundaries, add metadata features, and export feature collection as GEE asset
 - d) Combine the three feature collections into one, add a point ID for future merging, and export as GEE asset and as a shapefile for visualizations
- 4) Get Sentinel 2 band values from sampling points
- a) Load combined points collection and extract pixel values from Sentinel 2 images at the specified points
 - b) Export pixel data with date, location, and metadata to Google Drive (cannot export to asset due to GEE's memory limits even with this reduced dataset)
- 5) Get MODIS temperature values from sampling points
- a) Retrieve and process MODIS imagery collection
 - i) Apply processing functions that scale bands and set metadata properties
 - b) Extract pixel values from images at the specified points
 - c) Export pixel data with date, location, and metadata to Google Drive (cannot export to asset due to GEE's memory limits)
 - d) Extract temperature values from the single usable MODIS pixel in Provo Bay and export to Google Drive (not necessary for other waterbodies unless there is a similar issue with a small area entirely excluded by the 1km buffer)
- 6) Process extracted MODIS data
- a) Replace values for pixels located in Provo Bay with value of single unmixed pixel (not necessary for other waterbodies unless there is a similar issue with a small area where only a single MODIS pixel is valid)

- b) Replace values of pixels within 1 km of shore with nearest-neighbor values
- c) Impute missing data from partially-clouded images with daily lake median temperature
- d) Impute missing data from fully-clouded images with PCHIP temporal interpolation
- 7) Apply predefined chl-a and turbidity models to Sentinel 2 data
 - a) Load exported dataset of band values
 - b) Apply band models and filter for valid values
 - c) Impute missing data with PCHIP temporal interpolation
- 8) Combine and export final dataset
 - a) Merge processed Sentinel 2 and MODIS datasets and perform additional data cleaning and formatting.

2.5.3 Key outputs

- Waterbody boundary polygon shapefile.
- Random sample point dataset with coordinates.
- Chl-a and turbidity estimates for each point from Sentinel 2.
- Day and night surface temperature estimates for each point from MODIS.

This notebook provides example code that implements most of the processes we used to create the dataset published with this paper. It demonstrates data fusion between high-resolution optical (Sentinel-2) and low-resolution thermal (MODIS) imagery and generates daily data for analysis. The notebook is readily adaptable to other lakes where users can define their own sample regions or boxes, and provides a clear demonstration of how to generate valid, accessible, and useful water quality data from satellite imagery using the GEE platform.

2.6 User notes

We did not exclude data that appeared to be outliers—anomalous values due to sensor error or mixed pixels were already excluded by the QA process. We expect occasional very high and very low values in all three computed parameters because of the high variability intrinsic to environmental data.

There are no null or missing values in the dataset because any missing data (whether due to cloud cover, sensor error, or temporal gaps in the satellite imagery) were either imputed with the median value or interpolated. Imputed and interpolated data are flagged with a TRUE value in the int_flag column.

We generated the data from a spatial random sample, but these data do not meet the statistical assumption of independence (temporally or spatially) because they are next to each other in the same body of water and these processes have both spatial and temporal correlation.

Provo Bay is treated as a separate waterbody in Utah State law, so it may be useful to analyze it separately from the main lake using either the Boxes dataset or with the 'in_PB' data flag in the data, which identifies samples within Provo Bay.

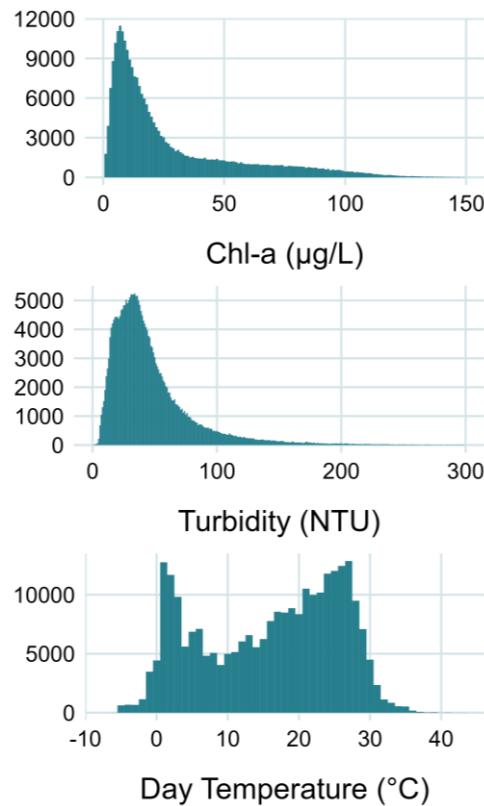


Figure 2.6: Histograms of the three parameters (combined datasets). Counts for temperature are higher because this is un-interpolated data, and MODIS has a daily revisit time, which more than doubles the amount of data compared to Sentinel 2 with its 2-3 day revisit time.

2.6.1 Summary statistics

We present summary statistics for the uninterpolated datasets except where otherwise specified. Since the purpose of the daily interpolated datasets is for time-series statistics that require regular time-steps, and the distributions of those datasets are highly similar to the uninterpolated datasets (see Figure 2.7), we did not provide summary statistics for the interpolated datasets other than Figure 2.7, which compares the distributions of the interpolated and uninterpolated datasets.

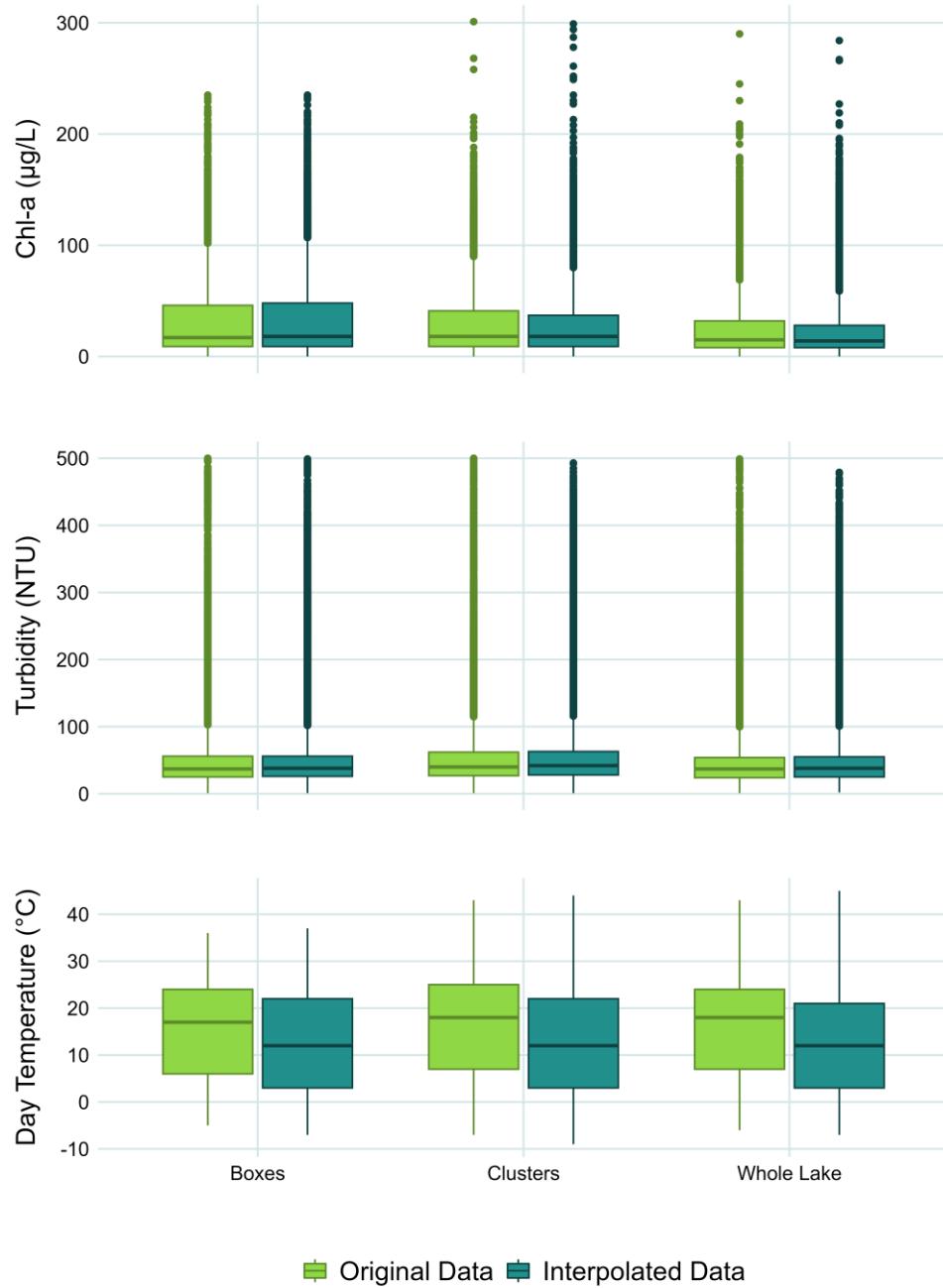
Table 2.2: Summary statistics for the three datasets. Note that we combined the subcategories within the Boxes and Clusters datasets for these descriptive statistics. Chl-a units are $\mu\text{g/L}$, Turbidity units are NTU, and Temperature units are $^{\circ}\text{C}$.

Dataset	Parameter	Min	Max	Standard Deviation	Interquartile Range	Skewness	Kurtosis
Whole Lake	Chl-a	0.0	290	24	21	1.86	6.30
Whole Lake	Turbidity	1	499	34	30	3.05	19.97
Whole Lake	Day Temp	-7	45	10	19	0.047	1.68
Whole Lake	Night Temp	-20	30	34	30	-0.12	1.92
Boxes	Chl-a	0	235	28	38	1.25	4.00
Boxes	Turbidity	1	500	35	30	3.49	25.20
Boxes	Day Temp	-7	37	10	19	0.05	1.70
Boxes	Night Temp	-18	27	9	16	-0.12	1.92
Clusters	Chl-a	0	301	26	29	1.53	5.05
Clusters	Turbidity	1	500	39	35	2.94	17.85
Clusters	Day Temp	-9	44	10	19	0.06	1.71
Clusters	Night Temp	-20	28	10	16	-0.10	1.93

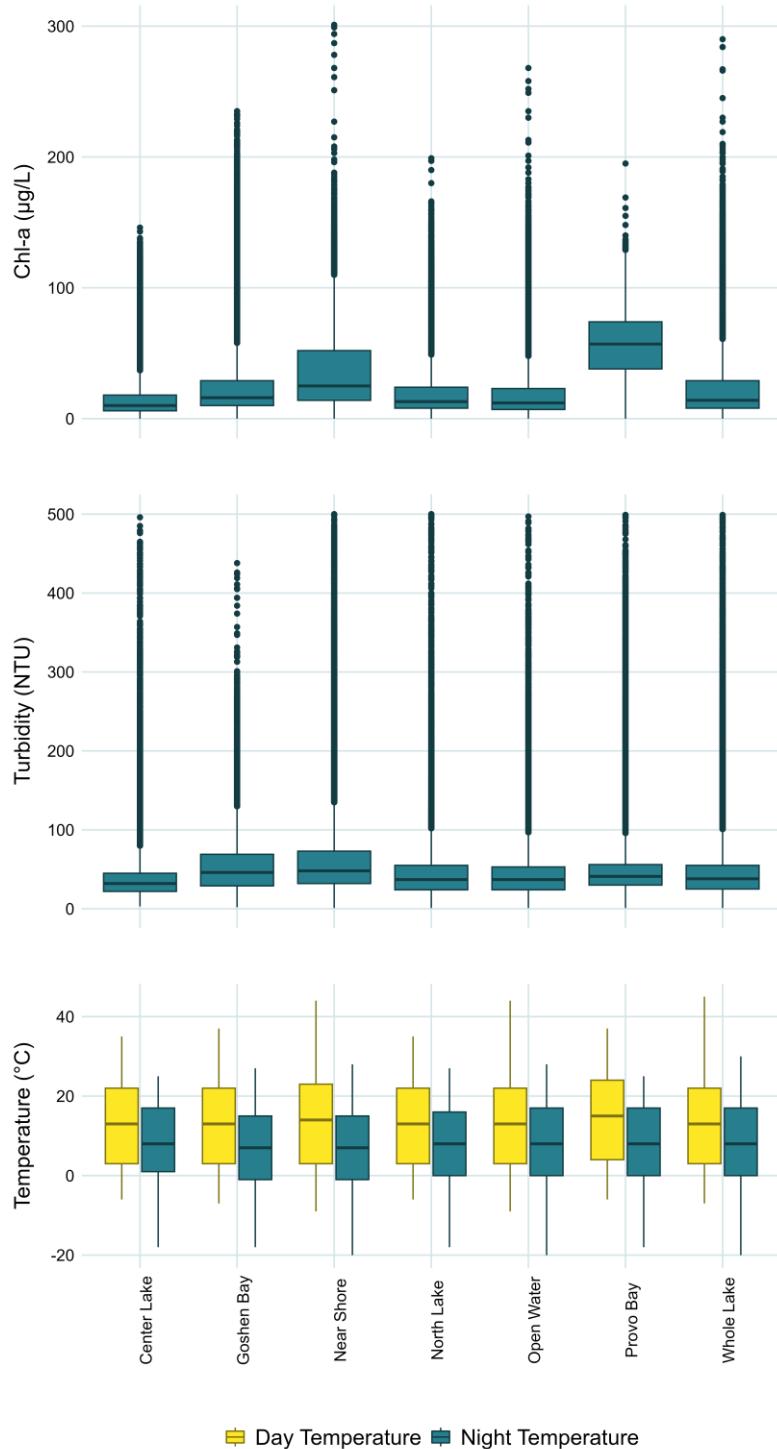
2.6.2 Summary plots

Below, we present several summary plots to show the range and characteristics of the data and demonstrate some of the insights gained by visualizing the dataset.

Figure 2.7 compares the distributions of the interpolated and original data (all datasets) for each parameter. The interpolated temperature data have a much tighter distribution than the measured data because the interpolated dataset is small. As discussed in Section 2.4, for Sentinel data, the numbers of PCHIP-interpolated points are 99%-120% the size of the number of measured points they were based on due to the longer gaps between Sentinel images, while the interpolated MODIS datasets are between 0.14% and 0.99% the size of the measured datasets, since MODIS has a daily revisit time and because missing points could be filled with median imputation rather than interpolation, except when the lake was completely clouded (which was rare). We provide Figures 2.6-2.12 to provide an overview of the data and its distributions. These graphs show both box plots and time series for the data grouped by location.



*Figure 2.7: Comparison of distributions of original and interpolated data. The box ends represent the 25th and 75th percentiles with the centerline showing the 50th percentile or median value. The whiskers represent 1.5*IQR (interquartile range) above the third quartile, with the dots representing the outliers, or values higher or lower than 1.5*IQR. The interpolated temperature distributions are much smaller because there are so few interpolated data points for temperature, as most missing temperature data points were filled with the median imputation method described in Section 2.4.2.*



*Figure 2.8: Distribution of chl-a, turbidity, and temperature measurements by sample set. The box ends represent the 25th and 75th percentiles with the centerline showing the 50th percentile or median value. The whiskers represent 1.5*IQR (interquartile range), with the dots representing outliers, or values higher or lower than 1.5*IQR above the third quartile.*

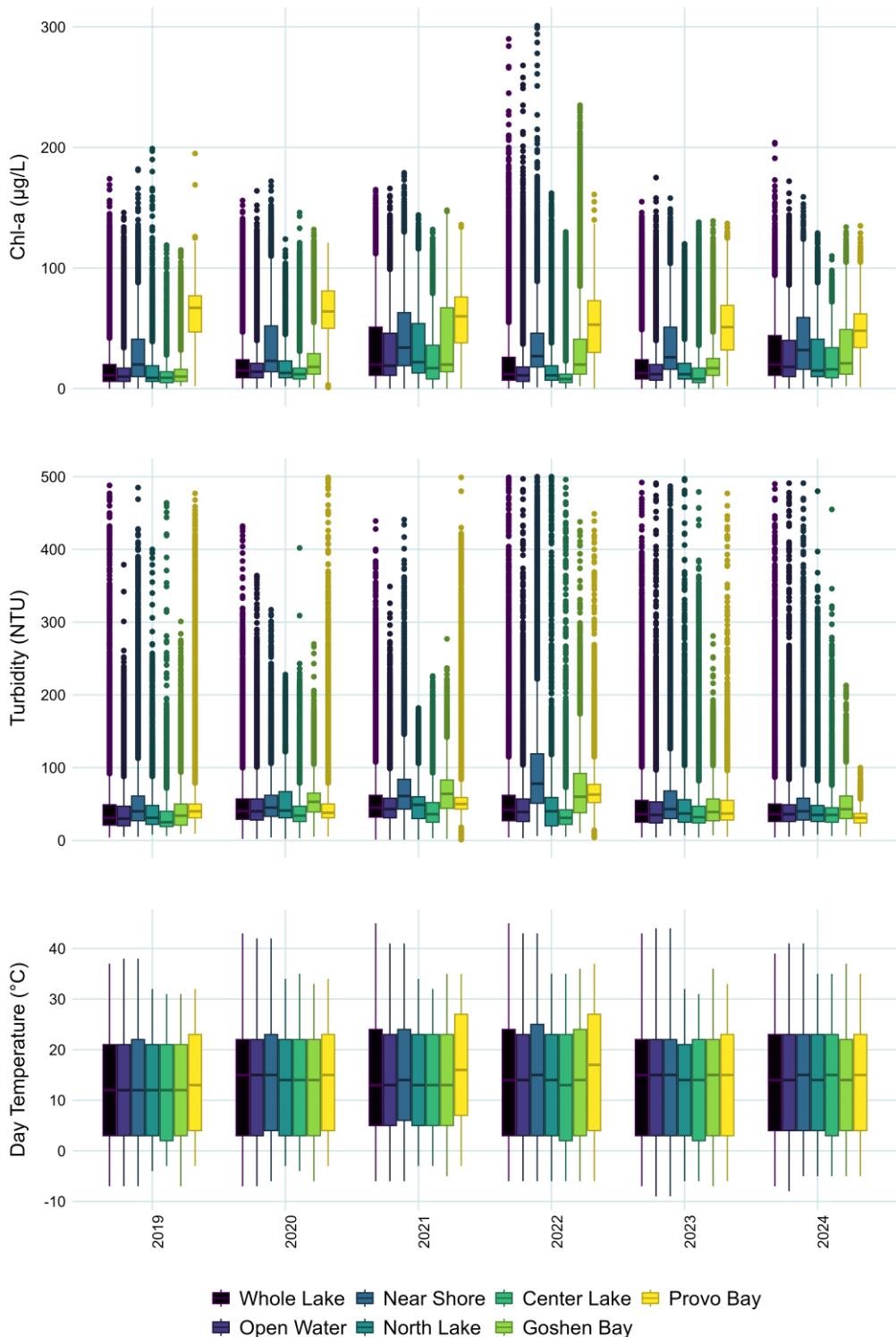
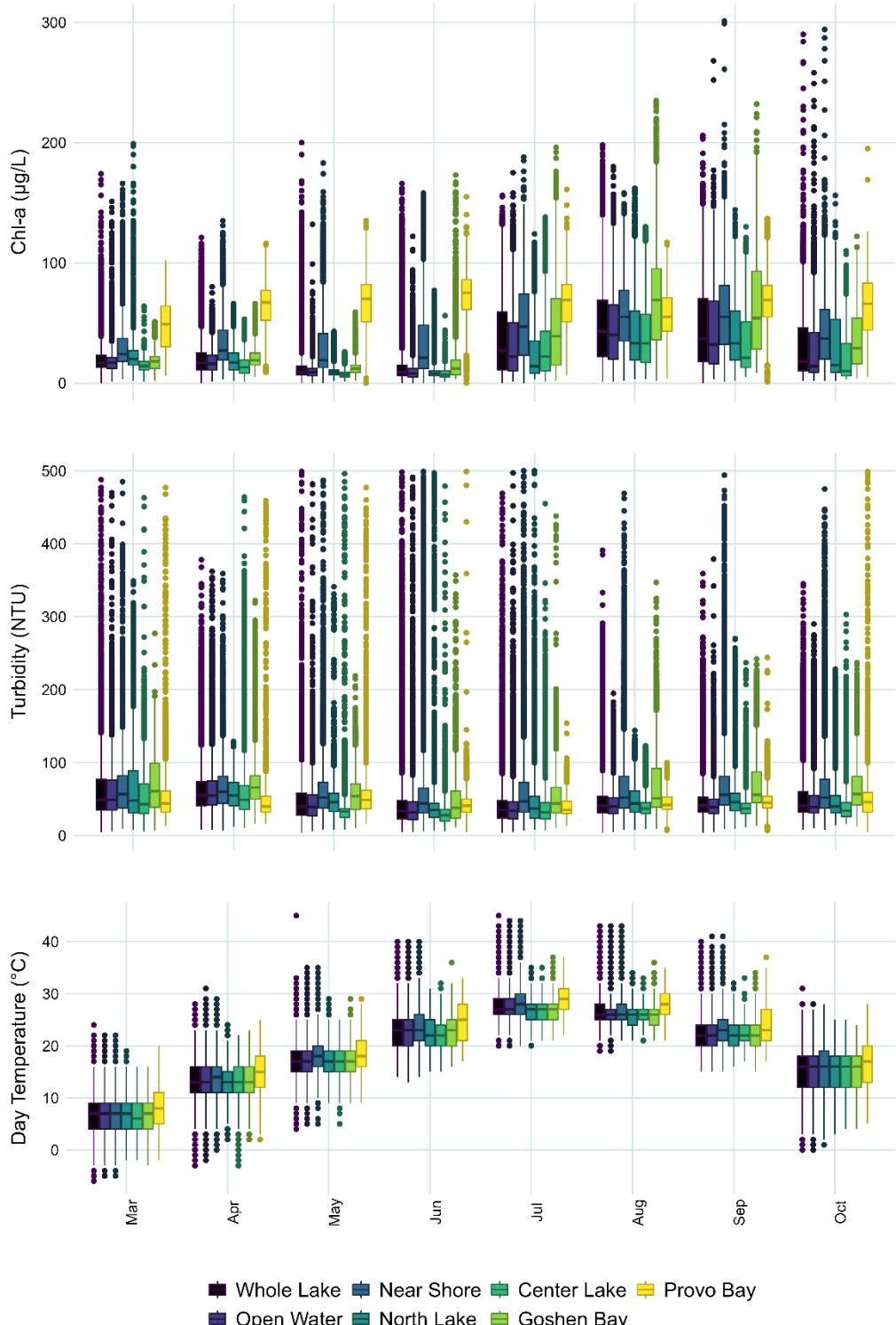


Figure 2.9: Distributions of the chl-a, turbidity, and temperature values by year and dataset. The box ends represent the 25th and 75th percentiles with the centerline showing the 50th percentile or median value. The whiskers represent 1.5*IQR (interquartile range), with the dots representing outliers, or values higher or lower than 1.5*IQR.



*Figure 2.10: Distributions of the chl-a, turbidity, and temperature values by month and dataset. The box ends represent the 25th and 75th percentiles with the centerline showing the 50th percentile or median value. The whiskers represent 1.5*IQR (interquartile range), with the dots representing outliers, or values higher or lower than 1.5*IQR.*

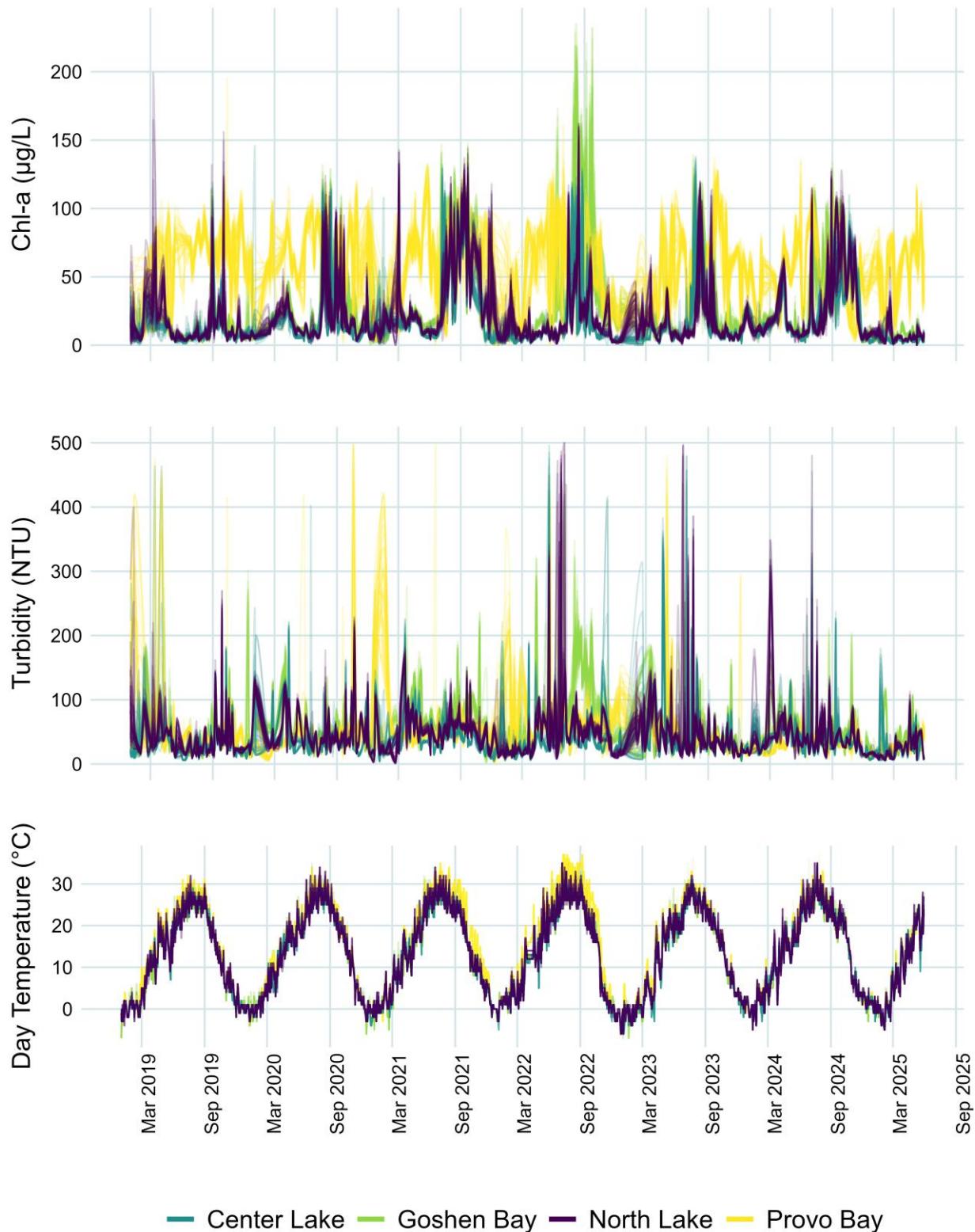
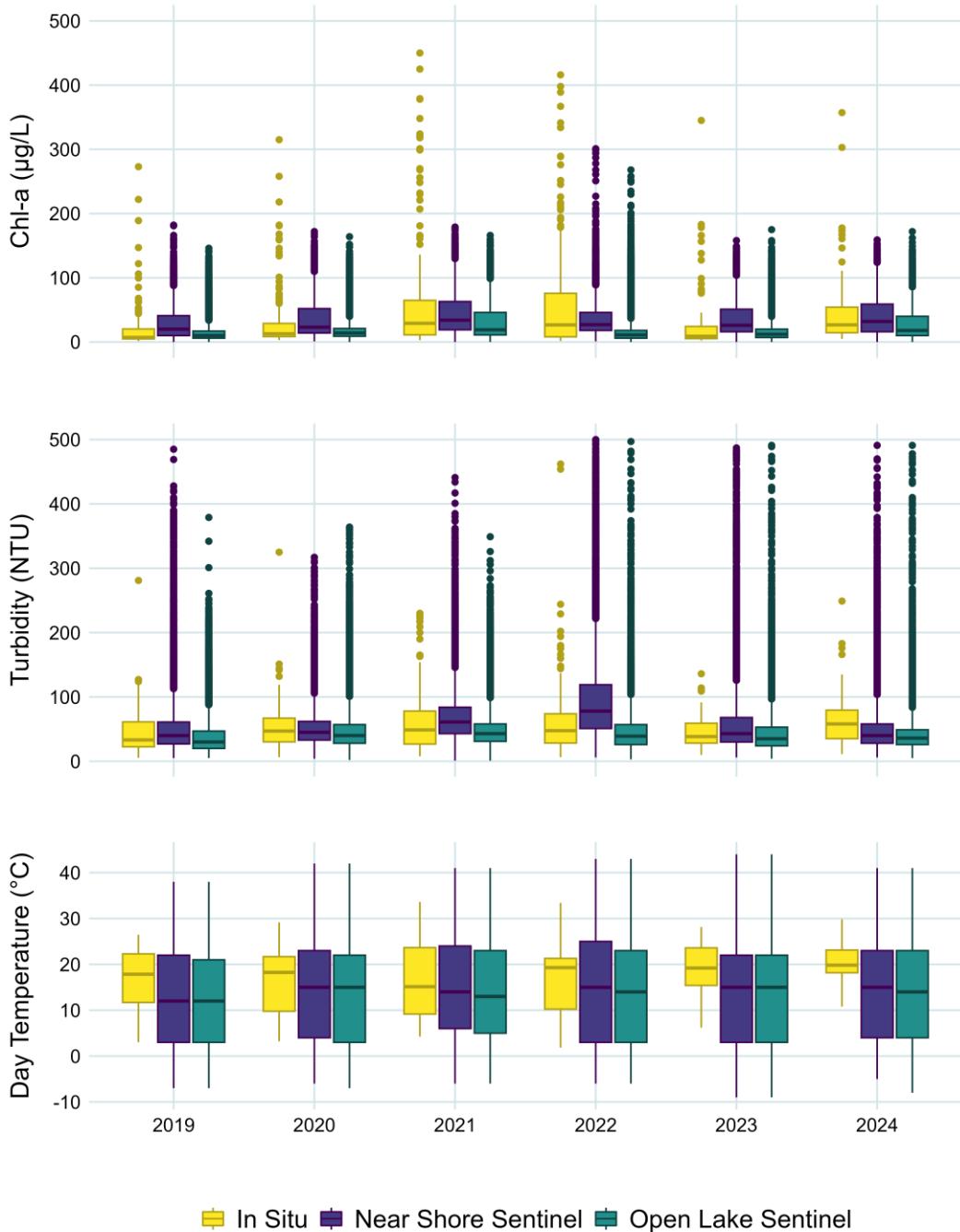


Figure 2.11: Temporal plots based on the interpolated boxes dataset to show the utility and validity of the daily interpolated data.



*Figure 2.12: A comparison of in situ data from the Department of Environmental Quality with our estimates of chl-a, turbidity, and temperature. The distributions match quite well with some differences, showing that the satellite data are valid and provide additional context about Utah Lake water quality missing from the spatially and temporally limited in situ samples. The box ends represent the 25th and 75th percentiles with the centerline showing the 50th percentile or median value. The whiskers represent 1.5*IQR (interquartile range), with the dots representing outliers, or values higher or lower than 1.5*IQR. We excluded four in situ chl-a measurements above 500 $\mu\text{g/L}$ from the plot for readability.*

Chapter 3: Correlation of chl-a with temperature and turbidity on Utah Lake

This chapter is based on the article *An earth observation data-driven investigation of algal blooms in Utah Lake: statistical analysis of the effects of turbidity and water temperature*, currently submitted for publication in MDPI Remote Sensing. In this chapter, we use the dataset described in Chapter 2 to analyze the relationship between chl-a, as an index for algal blooms, and the water column parameters of temperature and turbidity.

3.1 Introduction

Understanding how environmental conditions affect algal growth is essential for selecting bloom monitoring and mitigation strategies. Previous chapters provided an overview of current knowledge of algal blooms on Utah Lake and introduced how we use remote sensing data to generate comprehensive insights about the blooms. In this chapter, we use these data to analyze how spatial and temporal patterns of chl-a relate to two key physical conditions—turbidity and water temperature. We use various statistical and time series analyses applied to the satellite-derived datasets to extract information about the relationship between chl-a and these two variables.

Like nutrient concentrations, water temperature is a well-known driver of algal growth, but the influence of turbidity on algal growth is more complex. High turbidity usually has a negative effect on algal growth as it blocks sunlight from penetrating the water column [85-87] and can interfere with the bioavailability of

nutrients [88]. However, the relationship between algal growth and turbidity is complex because growth is also strongly influenced by nutrient availability [38] and higher turbidity can be associated with greater total amounts of nutrients in the water column [89]. In addition to this dichotomous effect, algal growth is strongly influenced by temperature [90,91]. Algal biomass itself also contributes to turbidity, further complicating a remote sensing analysis. These many factors necessitate advanced analysis methods to untangle the effects of turbidity on growth.

Experts on Utah Lake ecology have suggested that the lake's high levels of turbidity—caused by resuspension of soft benthic sediment, algal biomass, and carbonate precipitates—actually prevent more intense HABs by decreasing the available light and limiting algal growth in spite of high nutrient availability [92]. Light attenuation in the water column is a function of the size and concentration of suspended particles [89], and Utah Lake has very high amounts of suspended particles due to sediment resuspension caused by wave action and bottom-feeding fish. Sediment resuspension also maintains near-constant levels of nutrients in the water column since nutrient inputs are buffered by sorption processes with lakebed sediment [47]. Due to the high turbidity and relatively constant nutrient levels, algal growth in Utah Lake may be a light-limited system, where light availability, as governed by turbidity, and in combination with water temperature, determines the growth of algae. This assumption is supported by research on other shallow, turbid lakes which has found that turbidity does play a significant role in controlling algal growth [58,93], and by prior research on Utah Lake water column conditions [46].

It is well-established that HABs are stimulated by warm water, and that rising global temperatures are spurring more frequent and more intense blooms [4,94]. Both inter- and intraseasonal temperature variations influence algal growth [90], which is why temporally comprehensive research that quantifies seasonal and longer-term patterns is essential for understanding the relationships between algal growth, temperature, and turbidity. Some studies have found that temperature is a stronger predictor of algal growth than water column nutrient concentrations in

environments without nutrient constraints like Utah Lake [95]. Growing season water temperatures in Utah Lake range from 2°C (~36°F) in the early spring and late fall to 34°C (~93°F) at the peak of summer, with an average of 17°C (~63°F), based on data from the Utah State Department of Environmental Quality's Ambient Water Quality Monitoring System (AWQMS) database collected during the period covered by this study.

Because Utah Lake is a warm, turbid lake, quantifying the effects of these conditions on algal growth is essential for understanding bloom dynamics on the lake. Although nutrient levels are also an important driver of algal growth in many systems, they were only considered broadly in terms of distinct areas of the lake because they are not detectable by optical sensors such as satellites. Studies using data from the State's AWQMS database have shown that phosphorous water column concentrations have remained essentially constant, without discernible seasonal variation in the lake since the early 90s [47]. This study focuses on variation in algal growth associated with changes in temperature and turbidity, which, unlike nutrient concentrations, do fluctuate throughout the lake and the growing season.

To quantify these relationships, we employ generalized least squares (GLS) regression models to evaluate concurrent effects of temperature and turbidity on chl-a, coupled with analyses of outlier chl-a concentrations to identify conditions associated with intense algal blooms. We further examine temperature trends preceding the first bloom appearance each year to identify thermal triggers for bloom onset. By combining statistical models with temporal and spatial analyses, this chapter describes how turbidity-related light limitation and growing season water temperatures influence algal growth in Utah Lake.

This study focuses on two environmental factors known to strongly influence algal growth: temperature and light availability. Both of these parameters can be estimated on a spatially and temporally comprehensive scale with earth observation satellite data—water temperature found using emissivity measurements and light availability from optical turbidity measurements, since

there is a strong negative correlation between water column turbidity and light availability [41].

3.2 Analyses

3.2.1 Generalized least squares regression model

Algal growth is well-known to be strongly positively correlated with temperature in freshwater bodies worldwide [94], but the effects of turbidity on algal growth are less clear. To investigate the relationship between algal growth and turbidity in Utah Lake we used generalized least squares regression models that included daytime temperature as a covariate to identify whether there was any direct correlation between chl-a and turbidity measurements.

We used the nlme package in R to fit generalized least squares (GLS) models to each of the four groups of data in the Boxes dataset [96]. We focused on the Boxes dataset for this analysis because it allowed us to observe spatial variation in the results but was less computationally intensive than modeling the data from other two datasets, which had more data points in each category. We excluded interpolated and imputed values to reduce the computational load, since a consistent time-step was not necessary for this analysis. We included an autocorrelation structure in our models that assumed an exponential decay in the correlation between points in the time series to account for the fact that the datasets are comprised of repeated measures on sample locations and show strong temporal autocorrelation. Chl-a and turbidity values both follow an approximately log-normal distribution, and we found that using the log of both parameters improved the accuracy of the regression models.

For each category in the Boxes dataset (North Lake, Center Lake, Provo Bay, and Goshen Bay) we fit a “full” model that included both temperature and turbidity, and two “reduced” models—one with only temperature and the other with only turbidity. Additionally, in each of these models we accounted for residual temporal dependence. Specifically, let chl-a_{it} be the measurement of chl-a observed at location i and time t for a given Boxes category. Similarly, Temperature_{it} and Turbidity_{it} are the

measurements of day temperature and turbidity, respectively, at location i and time t . Note that due to the coarser spatial resolution of the MODIS imagery from which temperatures were obtained, for a given Boxes category, $\text{Temperature}_{it} = \text{Temperature}_t$ for all i . For each Boxes category, we fit the following reduced models,

$$\ln(\text{chl-a}_{it}) = \beta_0 + \beta_1 \text{Temperature}_{it} + \eta_{it} + \epsilon_{it}, \quad 3.1$$

$$\ln(\text{chl-a}_{it}) = \beta_0 + \beta_1 \ln(\text{Turbidity}_{it}) + \eta_{it} + \epsilon_{it}, \quad 3.2$$

where β_0 and β_1 represent covariate coefficients, η_{it} is a temporally-dependent random effect, and ϵ_{it} is an independent, zero-mean, normally-distributed error term with constant variance, σ^2 . All parameters were estimated uniquely for each model and Boxes category. The temporally-dependent random effect, η_{it} , is a zero-mean Gaussian process, with an exponential correlation structure such that,

$$\text{Cor}(\eta_{it}, \eta_{ju}) = \begin{cases} \exp\left\{\frac{|t-u|}{\phi}\right\}, & \text{when } i = j, \\ 0, & \text{when } i \neq j, \end{cases} \quad 3.3$$

and ϕ is a so-called range parameter estimated in the model, common across all locations in the Boxes category. The larger the value of ϕ , the higher the correlation will be for locations separated by a fixed time interval.

The full model is similar, but includes both covariates in the linear term. Specifically,

$$\ln(\text{chl-a}_{it}) = \beta_0 + \beta_1 \text{Temperature}_t + \beta_2 \ln(\text{Turbidity}_{it}) + \eta_{it} + \epsilon_{it}, \quad 3.4$$

where β_0 , β_1 , and β_2 are coefficients, η_{it} is a temporally-dependent random effect of the same form as in Equation 3.3, and ϵ_{it} is a zero-mean, independent normally-distributed error with constant variance, σ^2 . Temperature is the value of the MODIS estimate of daytime surface water temperature value, Turbidity is the NTU value estimated from Sentinel 2 data, and chl-a is the $\mu\text{g/L}$ value estimated from Sentinel 2 data. We used the lrtest function from the

lmtest package in R [97], which compares the suitability of nested models, to perform a likelihood ratio test on the full and reduced model for each category to investigate how well each variable is correlated with chl-a [98].

3.2.2 Outlier analysis of chl-a on turbidity levels

Chl-a and turbidity are highly variable both temporally and spatially on Utah Lake. Even with this variability, moderate to intense blooms are fairly rare relative to the lake's large size, high nutrient content, and long growing season—although blooms may affect isolated marinas and bays for several weeks during the summer, we almost never observe blooms covering a large area of the lake or lasting for more than a week in open water.

To investigate the relationship between turbidity and intense algal blooms, we focused on outlier chl-a values. We defined outliers as chl-a measurements above 87 µg/L, which was 1.5 times the interquartile range above the 75th percentile (third quartile) of the distribution of all chl-a values in the dataset (excluding the interpolated values) using the `is_outlier` function in R [99]. In the dataset, 6% of the values were above this threshold, and we assumed they represent moderate to intense algal blooms, distinct from background water column primary production and milder blooms.

We defined turbidity “bins” representing different levels of turbidity in the lake to analyze the behavior of chl-a values at varying levels of turbidity. The standard deviation of turbidity values in the entire dataset (excluding interpolated values) was 39 NTU, so we chose a bin size of 40 NTU for simplicity and interpretability. We analyzed the proportion of outlier chl-a values in each turbidity bin to investigate the relationship between turbidity levels and intense algal blooms.

There is a small but statistically significant seasonal trend in turbidity values in Utah Lake (Figure 3.1)—turbidity rises in the spring, then the water becomes clearer during the early summer before another peak in turbidity in the late summer and fall. Regression of turbidity values from the entire dataset using month as the explanatory variable yielded *p*-values below 0.001 for every

month, indicating that although the seasonal effect is relatively small (the range in monthly means is 25 NTU, while actual turbidity values in the dataset range from 1-500 NTU), it is a statistically significant variable that must be accounted for. Because of this we performed the turbidity and chl-a outlier analysis by month.

Investigating the cause of this seasonal cycle was beyond the scope of this work, but it may be explained by seasonal hydrologic changes. The peak in the spring may be due to heavy spring runoff flows washing sediment into the lake or resuspending lakebed sediment. The late summer peak could be associated with biological turbidity caused by algal blooms, or by low water levels that exacerbate the effect of wave-driven sediment resuspension. Additionally, since Utah Lake water is typically near the solubility limit of calcite, large evaporative losses (50% of the lake volume [100] in the late summer) may drive greater rates of calcite precipitation, which raises turbidity levels and gives the lake its characteristic milky color [7,100].

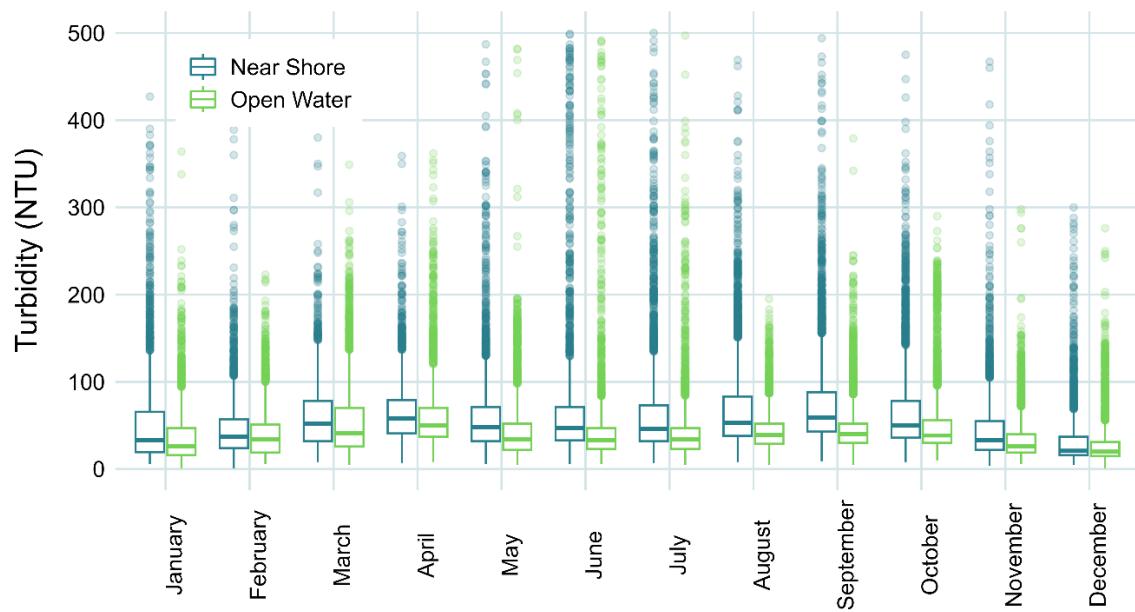


Figure 3.1: Turbidity boxplots by month showing seasonal trend. We used the clusters dataset (described in Section 2.4.3) to show that the trend is slightly different for near-shore areas.

We applied a Chi-square test to determine whether the difference in proportions of chl-a outliers between turbidity levels for each month and category was statistically significant. The Chi-square test requires a frequency of at least one per category (in this case, one outlier per turbidity bin and month), and that at least 80% of the expected frequencies are over 5. The expected frequency assumptions were violated for some of the turbidity bins and data categories, particularly at higher turbidity levels--many months had no outlier chl-a values in the turbidity bins above 200 NTU. To address this issue, we collapsed the nine turbidity bins over 200 NTU into a single bin after examination of the chl-a outlier frequencies showed that outliers very rarely occurred at these levels. We also used a Chi-square method in R that computes p -values using a Monte Carlo simulation, which allows the test to produce valid results even with low expected frequency counts. We calculated p -values for each combination of month and data category to account for the seasonal trend in turbidity values and to identify any spatial differences in the outlier proportions. Since the Chi-square test uses proportions of counts, it also accounts for the fact that lower turbidity levels are much more frequent in the dataset than higher ones.

3.2.3 Temporal temperature analysis

We created plots of temperature and chl-a data time series to visually assess the relationship between the two, and analyzed how changes in temperature affected the timing of higher chl-a values. Unlike the outlier analysis, which focused on more intense blooms, for this analysis we investigated the seasonal timing of the onset of “typical” primary production in the lake. The mean chl-a level during the growing season (which we defined as April through October) is 34 $\mu\text{g/L}$. So, for this analysis we defined a bloom as chl-a above 34 $\mu\text{g/L}$.

For each sample location in the Clusters dataset, we found the first instance of chl-a above 34 $\mu\text{g/L}$ each year and defined it as the first bloom. We first looked at the absolute temperature changes but found no discernable pattern, so we moved on to examining the first derivative of temperature. We calculated the daily rate of

change in temperature at each location for the 30 days preceding the first bloom to identify any patterns or correlations between first bloom appearance and temperature.

3.3 Results

3.3.1 Loose correlation of chl-a with turbidity varies temporally and spatially

As can be seen in Figure 3.2, which contains pairwise plots of the natural log of the turbidity and chl-a values by lake area and month, the direct correlation between chl-a and turbidity measurements is complex and varies greatly in different areas of the lake. For additional plots of chl-a and turbidity separated by year, see Appendix A.

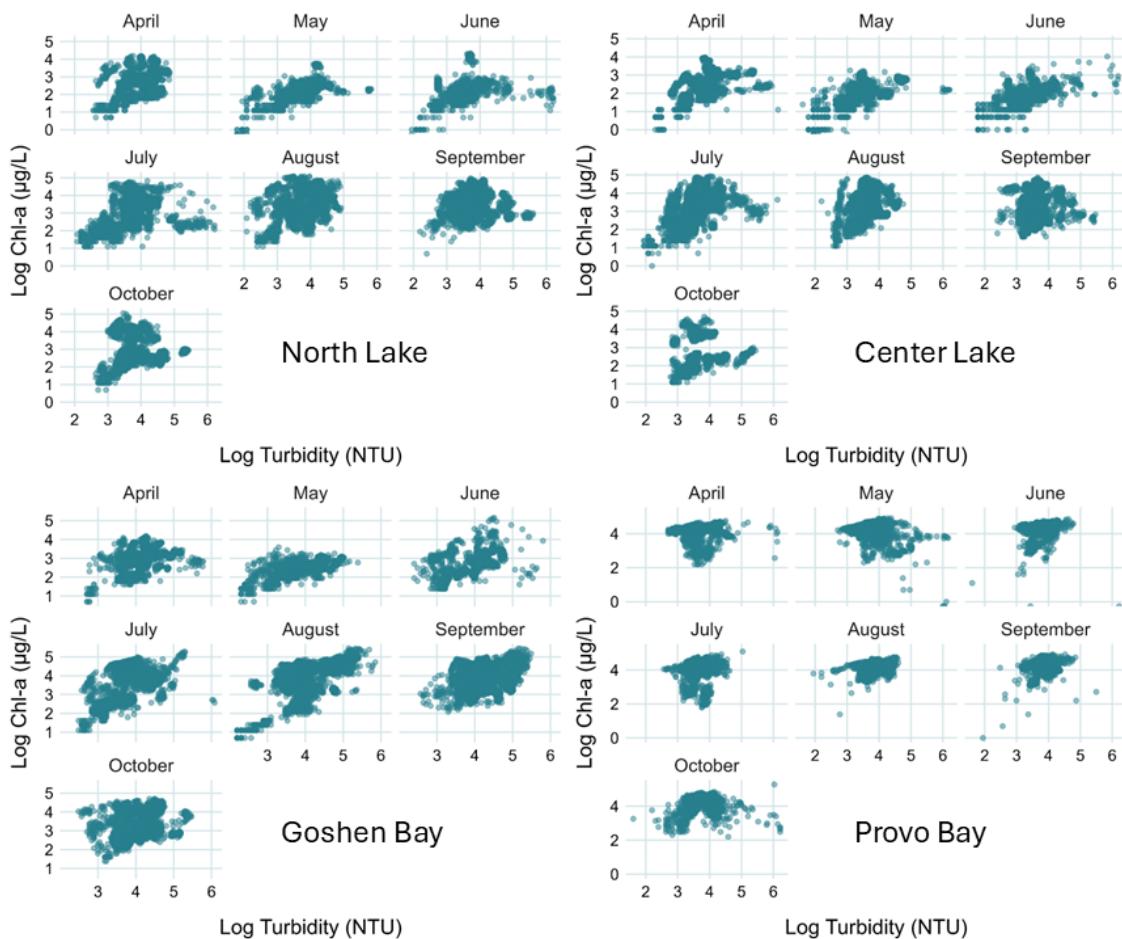


Figure 3.2: Chl-a vs. turbidity measurements for the Boxes dataset.

In the data for Goshen Bay, the relationship between chl-a and turbidity appears to be generally positive in July, August, and September, but uncorrelated for the rest of the months and for most months in the other datasets. The right skew present in most months is at least partially due to the higher frequency of low-turbidity measurements on the lake, but it is noteworthy that high chl-a values typically do not occur with high turbidity values except in Goshen Bay in the months of July, August, and September. The higher frequency of lower turbidity values in the dataset also complicates any correlation. In addition to the complexity caused by the various environmental factors beyond turbidity that affect algal growth, the nature of turbidity measurements also makes it more difficult to identify the true relationship between chl-a and turbidity: the presence of algal biomass increases turbidity levels, so even if high turbidity suppresses algal growth, the relationship may appear to go in the opposite direction due to the effects of algal growth on the turbidity measurements themselves.

The GLS models were unable to predict chl-a with any reasonable accuracy using only temperature and turbidity values, which is not surprising given the complexity of the factors driving algal growth, but the coefficients for the full and reduced models suggest some interesting interactions and processes.

Table 3.1: Coefficients for full models. Note that the model is predicting the natural log of the chl-a values.

Category	Coefficient (Standard Error)		
	Temperature	ln(Turbidity)	Temporal Range
Provo Bay	0.024 (0.001)	-0.185 (0.010)	5.47
Center Lake	0.027 (0.001)	0.425 (0.006)	9.7
Goshen Bay	0.032 (0.001)	0.318 (0.006)	10.58
North Lake	0.018 (0.001)	0.270 (0.006)	8.78

All four models selected a small positive coefficient (between 0.02 and 0.03) for temperature, which was expected, although the small coefficient relative to the range of the temperature

measurements suggests that temperature explains only a small part of the variation in chl-a values in Utah Lake. The coefficients having close to the same value in all four locations, however, suggest that the effect of temperature is fairly constant throughout the lake.

The models for Center Lake, Goshen Bay, and North Lake selected small positive coefficients (0.42, 0.32, and 0.27, respectively) for turbidity, and the Provo Bay model selected a small negative coefficient (-0.18), suggesting that overall turbidity values are also not well-correlated with chl-a values. Unlike temperature, however, the effect varies spatially throughout the lake, particularly in Provo Bay, where the relationship is opposite to that of the other three areas.

We compared the Akaike's Information Criteria (AIC) statistics for the reduced models with turbidity only and temperature only to assess whether turbidity or temperature explained more of the variation in chl-a values. As shown in

Table 3.2, the models based on turbidity had lower AIC scores for all locations except Provo Bay, meaning they predicted more of the variation in chl-a values than the models based on temperature.

Table 3.2: Comparison of AIC values for reduced models

	Temperature Model	Turbidity Model	Difference
Provo Bay	22105	23064	-959
Center Lake	48783	44898	3885
North Lake	46783	45242	1541
Goshen Bay	36444	35065	1379

Although turbidity is a better predictor of chl-a than temperature by itself, as indicated by lower p-values for the turbidity coefficients and the AIC scores, the likelihood ratio tests yielded significant p -values ($p < 0.001$) and large Chi-square values for every model, indicating that the full models including temperature were better at predicting chl-a than the reduced models with either turbidity or temperature alone. The model improvement was greatest for the Provo Bay location, with a Chi-

square value of 1298.4, compared to 943.86, 713.83, and 330.67 for Goshen Bay, Center Lake, and North Lake, respectively.

3.3.2 Chl-a outliers associated with low turbidity values

Although linear models based on turbidity and temperature were not good predictors for chl-a values and found only a slight positive correlation between chl-a and turbidity overall, there was a negative relationship between turbidity and unusually high chl-a values. There is a statistically significant association between the relative proportion of outliers in a turbidity bin and the turbidity level of that bin. The three lowest turbidity bins, [0, 40), [40, 80), and [80, 120), contained 4,969, 8,952, and 1,365 of the chl-a outliers, respectively, which, combined, is 94% of the total number of outliers. The rest of the bins combined (10 bins) contained 1,042 outliers. This effect is partially due to the relative rarity of measurements in the higher turbidity categories, especially in Provo Bay.

Although turbidity levels above 400 NTU were occasionally observed on the lake, these represent fairly rare events of very extreme turbidity. Most “healthy” freshwater lakes are typically around 10 NTU and levels above 100 NTU are considered unsafe for most aquatic life. Figure 3.3 shows the proportion of chl-a outliers for each turbidity bin colored by month. Although the number of outliers was calculated based on the distribution of the entire dataset, the total counts were separated by month for this plot to show the seasonal trend. For many months, there is a higher proportion of chl-a outliers in the lower turbidity bins, especially in Provo Bay.

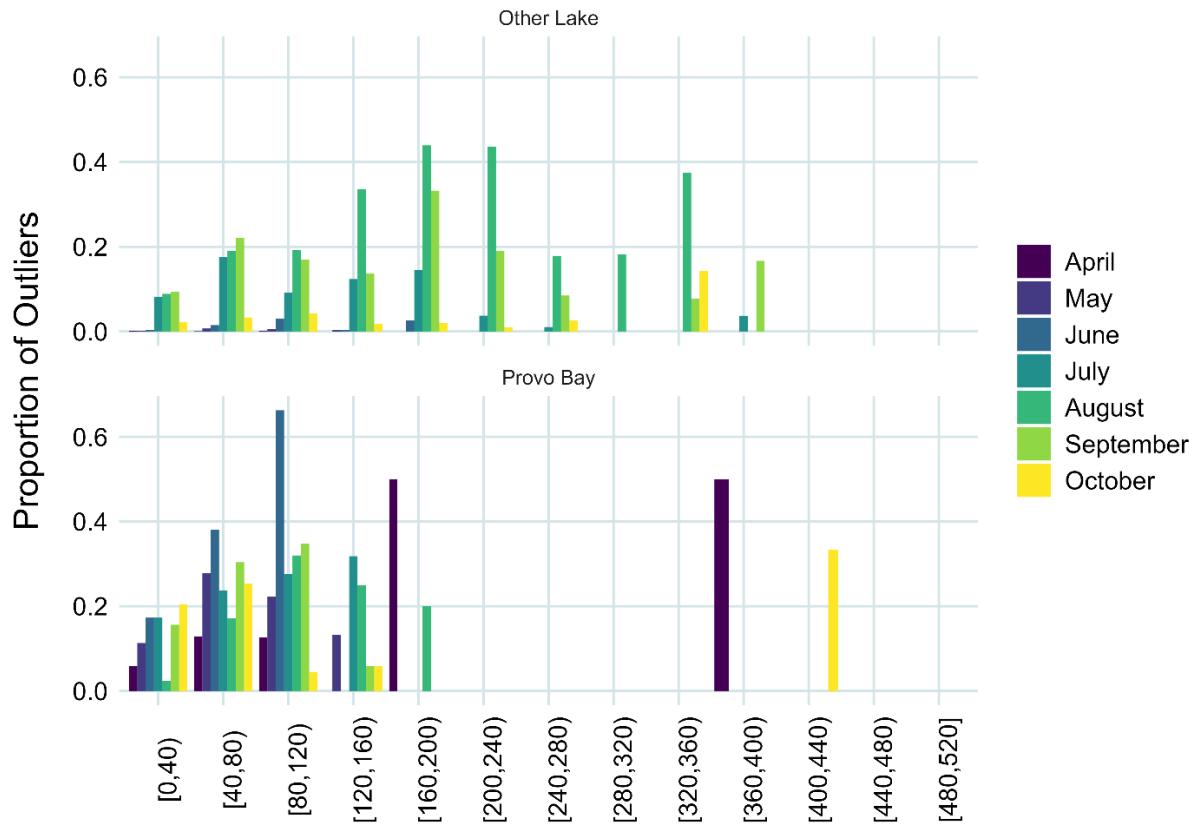


Figure 3.3: Chl-a outlier proportions by turbidity bin. Note that outliers were defined using the entire dataset while the counts are grouped here by month to show seasonal effects. Outlier values ranged from 87-301 µg/L and represented 6% of the satellite-derived chl-a values.

As shown in Figure 3.4, the Chi-square test for proportions yielded a significant p -value (<0.05) for the majority of the tested categories and months. A few of the categories did not have a sufficient number of outliers to allow the test to run: specifically categories in the Boxes dataset—which had just 50 sample points per category vs. 100 or 200—in the earlier months when chl-a values are typically lower.

	April	May	June	July	August	September	October
Goshen Bay	NA	NA	0.02670	0.00010	0.00010	0.00010	0.00010
Provo Bay	0.00140	0.00010	0.00010	0.00030	0.00010	0.00010	0.19228
Center Lake	NA	NA	NA	0.00010	0.00010	0.08439	0.00470
North Lake	NA	NA	NA	0.89881	0.00010	0.00010	0.00200
Near Shore	0.09049	0.00010	0.00240	0.00010	0.00010	0.00010	0.00010
Open Water	NA	1.00000	0.27777	0.00010	0.00010	0.00010	0.81812
Whole Lake	0.34447	0.01530	0.00030	0.00010	0.00010	0.00010	0.11429

Figure 3.4: Chi-square test results for dataset categories and months. Brighter green shows categories with a significant p-value result, indicating that the differences between the proportions of outliers in each turbidity bin for that month and location were not due to random chance. Grey values represent months when there were not enough outliers in that location dataset for the test to work. Notice that only the location datasets which include the area of Provo Bay (Provo Bay, Near Shore, and Whole Lake) had a sufficient number of outliers during the early months for the test to run, confirming previous observations that algal blooms begin in Provo Bay earlier in the season compared to the rest of the lake.

3.3.3 First algal bloom often preceded by a large temperature increase

Time series plots of chl-a and temperature graphs by year (Figure 3.5) show several interesting inter- and intraseasonal patterns. First, many near-shore locations experienced high chl-a levels much earlier in the year than open-water locations in every year, though by July there was typically no meaningful difference in chl-a concentrations between open-water and near-shore locations. There was also a small peak in chl-a concentrations early in the spring followed by a period of lower activity before the high concentrations of late summer and early fall. Chl-a concentrations also remain high in the fall even after temperatures drop to or below early spring levels.

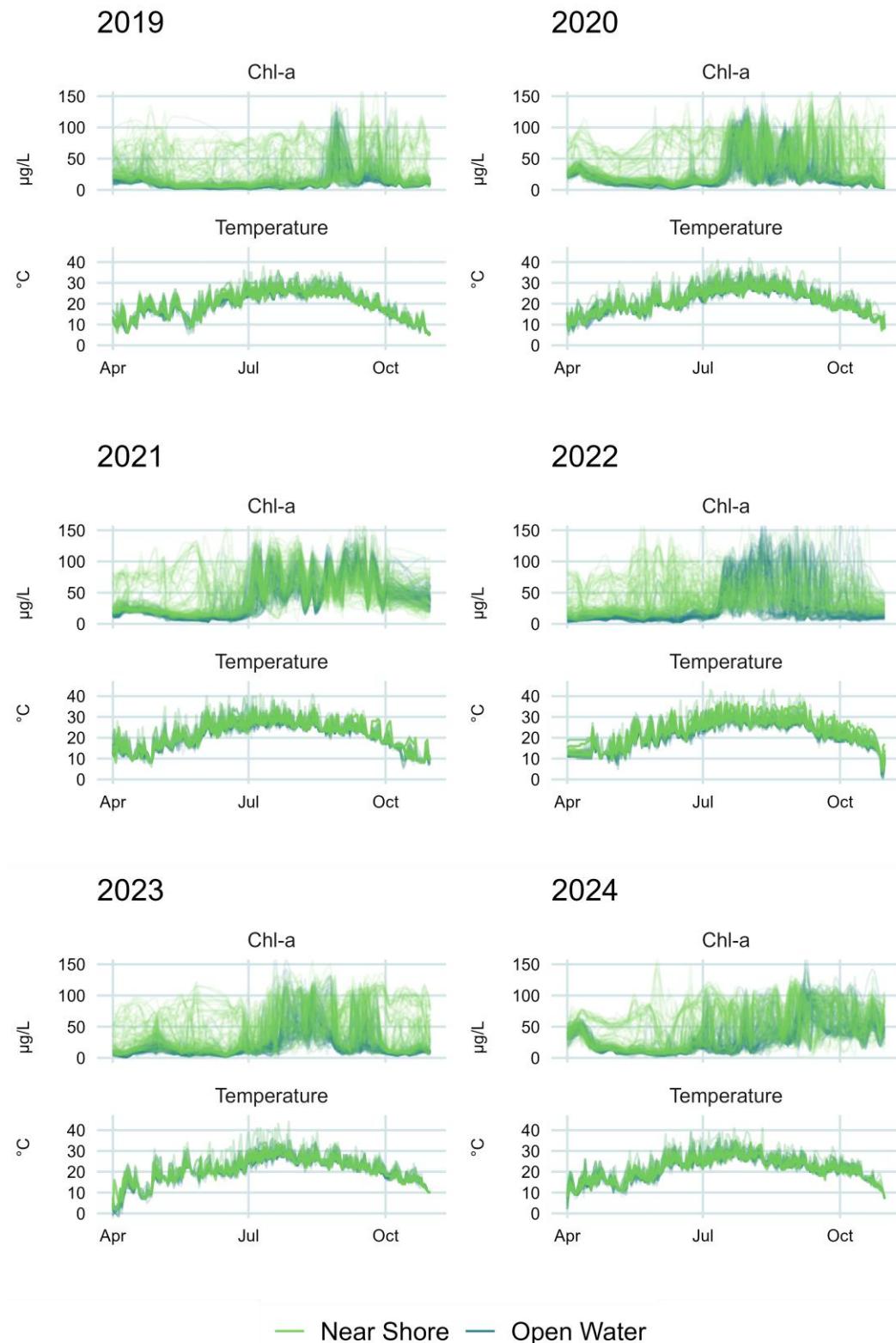


Figure 3.5: Comparison of chl-a and temperature values by year. Each line represents a single location with 600 locations in each plot.

We found no meaningful patterns associating intense blooms (above 87 µg/L) with temperature—either specific temperature thresholds or temperature changes. However, the analysis of milder blooms, defined as chl-a concentrations above the growing season median of 34 µg/L, showed some potential temperature-driven effects. Specifically, we observed that when the first bloom of the season occurred during the month of April, which it usually did for near-shore areas, it was often preceded by a larger-than-normal increase in temperature relative to the previous 30 days. Figure 3.6 shows temperature values for each location in the Clusters dataset by year for the 30 days leading up to the first occurrence of chl-a over 34 µg/L (mild bloom) at that location for that year. The lines are colored by the month in which the first mild bloom occurred. It is apparent that almost all near-shore locations experience a mild bloom during the month of April when temperatures are lower (purple lines at the bottom of each graph), while many open-water locations do not exceed the 34 µg/L threshold until June or later, when temperatures are higher (blue and green lines at the tops of the graphs). This was not the case, however, in 2024, when every sampled location on the lake crossed the 34 µg/L threshold in April.

We found the daily rate of change in temperature values at each location for the 30 days preceding the first bloom of the season. We defined the maximum temperature change as the highest positive daily rate of change which occurred at that location during the 30-day period. The maximum temperature change was not defined by a specific threshold, rather it was the largest positive temperature change that occurred at that location before the first mild bloom of the season. Maximum temperature change values ranged from 3-12 °C for all locations.

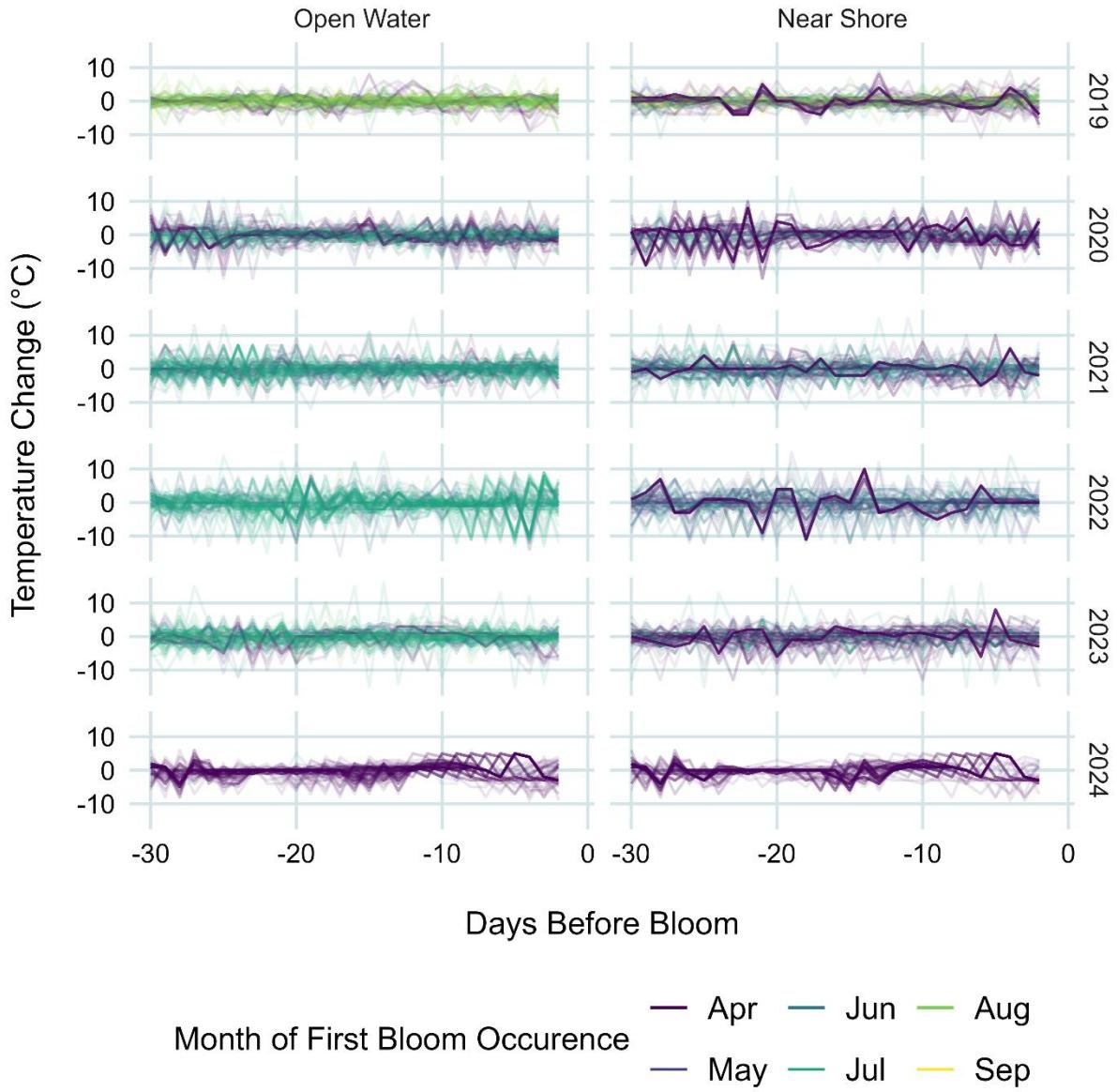


Figure 3.6: Daily temperature change preceding the first occurrence of chl-a over the mean ($34 \mu\text{g/L}$), colored by the month of the occurrence. For both open-water and near-shore locations there is no distinct temperature pattern for late-season blooms, but when chl-a over $34 \mu\text{g/L}$ first occurs in April it is typically preceded by a sharp increase in temperature between 3 and 10 days prior. This graph also shows that near-shore locations typically experience chl-a levels above the mean in April, while open-water locations rarely saw concentrations above the mean until later in the summer (except in 2020 and 2024).

Figure 3.7 shows histograms of the occurrence of maximum temperature change values by the number of days before the bloom happened. The top plot shows data for instances when the first bloom at a location occurred during the month of April. The bottom

plot shows data for instances when the first bloom at a location occurred during any month after April. As can be seen in Figure 3.7, blooms occurred 5 days after the maximum temperature change at that location in a disproportionate number of instances: 148 of the maximum temperature changes occurred 5 days before a bloom, while all other days had an average of 23 occurrences of the maximum temperature change. As shown in the bottom panel of Figure 3.7, which shows instances when the first bloom occurred in any month after April, the pattern exists only when the first mild bloom of the season occurs in April. The effect does not appear to be due to temperatures crossing a specific threshold as the water warms, since temperatures on the days of first bloom occurrence ranged from 7-28 °C (45-83 °F) for all locations and years.

3.4 Discussion

We analyzed a six-year dataset derived from Sentinel-2 and MODIS satellite imagery to examine relationships between chl-a, turbidity, and surface water temperature in Utah Lake. The analyses revealed that while direct correlations between turbidity, temperature, and chl-a are weak and spatially and temporally variable, the probability of high chl-a concentrations—indicative of intense algal blooms—is associated with low turbidity conditions, and the first mild bloom of the season often occurs a few days after an unusually large temperature increase relative to temperatures over the previous 30 days at that location. This research builds on previous studies which used data from the Landsat missions to analyze long-term trends in algal blooms [13,59] by incorporating higher-frequency data from Sentinel 2 and MODIS to investigate correlations between algal blooms and water column conditions.

3.4.1 Overall correlation of chl-a with turbidity

The results of the GLS models showed a small positive correlation of chl-a with turbidity values for all locations except Provo Bay, where the trend reversed—likely representing the effect of algal biomass increasing water column turbidity levels. A comparison of simplified models based on either temperature or turbidity alone

showed that turbidity is better correlated with chl-a values compared to temperature alone for all locations except Provo Bay; however, there was a small but statistically significant improvement in predictive power when temperature was included as a covariate in the model alongside turbidity for all locations. The turbidity coefficients for the model predicting the natural log of chl-a values ranged from -0.18 to 0.42, while temperature coefficients were between 0.2 and 0.3. The results of the GLS model suggest weak linear associations, consistent with the hypothesis that multiple, interacting processes influence chl-a dynamics.

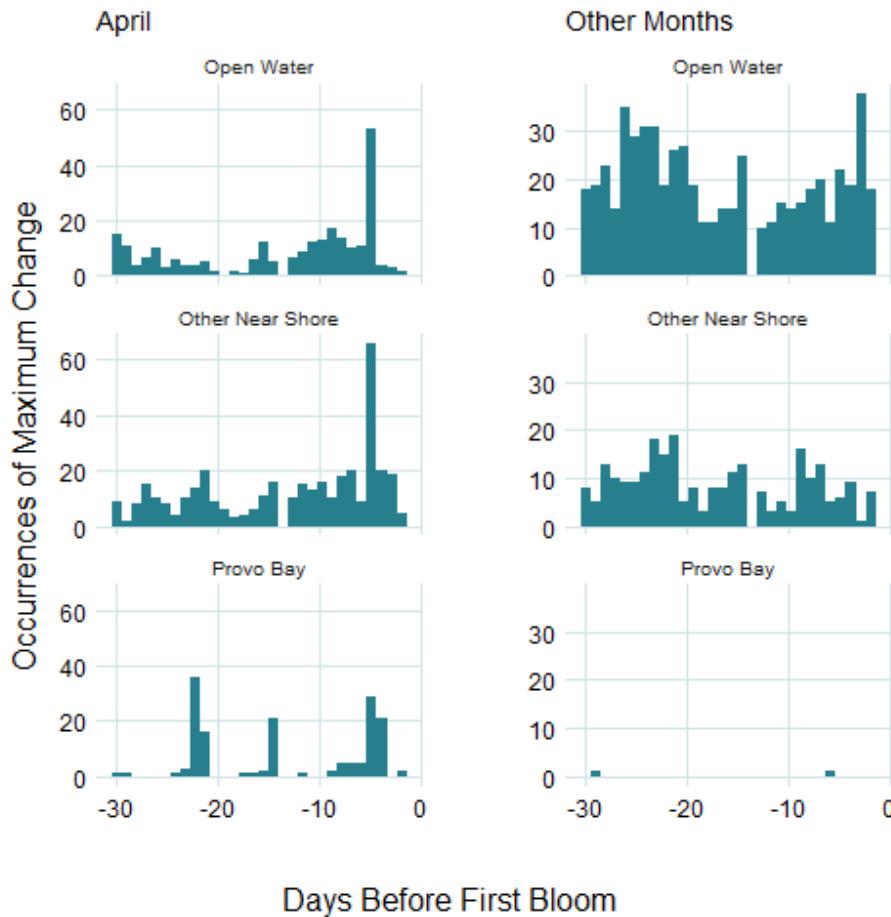


Figure 3.7: Histograms of occurrences of maximum change in temperature during the 30 days preceding the first bloom of the year at a given point. There is a large spike at 5 days before the bloom (148 occurrences as opposed to an average of 23 occurrences for the other days) for blooms that occur in April. For blooms that occur in all other months there is no distinct pattern.

The temporal range for the full model of the data from the Provo Bay category was much smaller than the temporal ranges of the full models for the other three categories, indicating that chl-a levels are correlated at longer temporal intervals in the main lake and Goshen Bay than in Provo Bay. This suggests that blooms in Provo Bay experience greater temporal variability than blooms on the main lake and in Goshen Bay. As the distributions of the data from each category show that chl-a concentrations in Provo Bay are typically higher than in the other areas of the lake, this could be due to algal growth remaining at “background” levels in the main lake during the growing season, while Provo Bay experiences blooms of varying intensities.

The significance tests of the models suggest a very loose overall positive correlation of chl-a with temperature in all areas, and with turbidity in all areas except Provo Bay. This positive correlation is likely the effect of algal biomass on turbidity measurements. The reversed relationship in Provo Bay may be due to the much clearer water conditions that dominate there, causing the effect of biological turbidity to be overcome by the greater light availability increasing algal growth.

A recent in situ study of algal response to nutrients on Utah Lake by Lawson, *et. al.* [37] found that, in small mesocosms located at the lake surface in a marina on Utah Lake, certain algal species were responsive to nutrient additions. Studies on other shallow, turbid, eutrophic lakes have similarly found that even when prevailing nutrient concentrations are very high, variations can still affect algal growth [38,39]. The relatively weak predictive power of turbidity and temperature for chl-a concentrations in our models generally suggests that other factors are also important for algal growth—most likely nutrient concentrations and other variables including water circulation patterns and other geochemical conditions. But, crucially, it appears that light availability may modulate to what extent the algal blooms can use available nutrients [39,101].

Interestingly, Liu *et. al.* [40] found that high turbidity suppressed diatoms and green algae and favored dominance of cyanobacteria in a large, shallow lake similar to Utah Lake. Our data did not

differentiate between chl-a from green algae, diatoms, and cyanobacteria, because the spectral resolution of the Sentinel 2 sensors is not sufficient for differentiating between them, and if green algae and cyanobacteria respond differently to light and turbidity this could partially explain the poor predictive power of our models, which predicted a combined value representing both types.

3.4.2 Correlation of outlier chl-a values with turbidity

In our data, the effects of light limitation on algal growth were mainly apparent for intense blooms. Chl-a values above 87 µg/L, which we defined as outliers based on the rule of 1.5 times the IQR above the third quartile, were assumed to represent intense bloom events. Just 6% of the chl-a values in the dataset were outliers, and they occurred disproportionately at turbidity levels below 120 NTU. Specifically, 15,286 outliers occurred in the three lowest turbidity bins (0–120 NTU, bin width = 40 NTU), while 1,042 occurred in the 10 bins above 120 NTU combined. Chi-square tests applied by month and region confirmed that this pattern was statistically significant ($p < 0.001$) across most categories, especially in near-shore areas and Provo Bay. This implies that intense algal blooms occur more often when turbidity levels are low, an effect likely due to light-limitation of algae growth caused by high turbidity. This pattern suggests that turbidity-induced light-limitation may be an inhibitor of severe algal blooms on Utah Lake and should be taken into account when assessing the potential effects of any proposed mitigations strategies. Yip, *et. al.* [101] found that water clarity and chl-a were positively correlated on a shallow lake similar to Utah Lake, which supports this finding.

A 2018 report by the Utah Department of Environmental Quality states that Utah Lake is not light-limited based on findings that overall levels of chl-a in Utah Lake are within the expected range for the overall levels of total phosphorus in Utah Lake [102]. Our analysis shows that these overall levels based on in situ measurements may not fully characterize algal bloom drivers on Utah Lake, as the remote sensing data show that intense blooms often occur when and where turbidity levels in the lake are lowest.

Despite high nutrient concentrations that support excessive algal growth, intense bloom formation may be constrained by the light-limiting effects of suspended sediments. High turbidity reduces underwater irradiance, thereby inhibiting photosynthesis even under favorable thermal and nutrient conditions. This finding fits well with other research on shallow lakes, both *in situ* studies [41,89,93] and a study similar to ours using Landsat [101] which found that light availability affects the extent to which algae are able to make use of available nutrient concentrations and supports the hypothesis that light limitation acts as a regulating mechanism for algal blooms in Utah Lake.

3.4.3 Correlation of chl-a with temperature

Temperature exerted a consistent positive influence on chl-a, in line with the consensus understanding of algal bloom dynamics [1,85,90]. However, the relatively small temperature coefficients in the GLS models suggest that thermal effects, while relevant, are less predictive of bloom occurrence than turbidity-induced changes in light availability. The results highlight the importance of integrating light availability into bloom risk assessments, particularly for shallow systems with highly variable turbidity.

In addition, temperature changes, rather than specific temperature levels, seem to have a more significant effect on the timing of the first mild bloom of the season. A disproportionate number of first instances of chl-a over the mean ($34 \mu\text{g/L}$) occurred five days after the largest positive temperature change in 30 days at that location. Since the absolute temperature values at the time of the first bloom ranged from $7\text{-}28^\circ\text{C}$ ($45\text{-}83^\circ\text{F}$), and the GLS models showed only a weak overall correlation with temperature, this suggests that an unusually hot day may be more likely to trigger algal growth than more gradual temperature increases or prolonged periods of hot weather. This effect was only present when the first bloom occurred during the month of April, however, suggesting that other factors (including turbidity) dominate later in the season.

Spatial analysis indicated that Provo Bay exhibits earlier and more frequent bloom activity than other regions of the lake, which

is consistent with previous studies [59]. It was the only subregion with sufficient chl-a outliers to perform statistical tests in early spring months (i.e., April), suggesting an earlier bloom onset. The time series graphs of chl-a show many near-shore locations (which include all of Provo Bay) with high chl-a levels in April, while other areas of the lake do not see similar levels until June or later. This may be caused by the bay's hydrologic isolation, shallow depth, and inflows with elevated nutrient concentrations. Provo Bay is also typically much less turbid than the rest of the lake, so the fact that blooms arise earlier in this area lends support to the finding that blooms are associated with less turbid water.

3.5 Conclusions

Overall, these results underscore the complex, nonlinear interactions between turbidity, temperature, and algal biomass in Utah Lake. Remote sensing data cannot resolve all causal mechanisms, and estimates of turbidity, temperature, and algal biomass derived from remote sensing data always involve some error, but despite these limitations, remote sensing enables high-resolution spatial and temporal analysis of conditions associated with bloom development. The findings presented here provide evidence that turbidity-related light limitation may be a constraint on intense algal bloom formation in this system and should be explicitly considered in future management strategies and modeling efforts.

The results also suggest that the first mild blooms of the growing season occur within days of a spike in water temperature, which monitoring efforts could consider to improve timely detection of blooms. Although nutrient levels are an important driver, the effects of light limitation and water temperatures on algal growth in Utah Lake cannot be ignored.

Only 6% of the chl-a measurements collected over a six-year period qualified as outliers in the distribution of the entire dataset, with a value over 87 µg/L. For this study, we assumed that these outliers indicate the presence of a bloom. This supports our personal observations from extensive field work on Utah Lake and

previous research showing that blooms rarely cover large areas of the lake outside of isolated marinas and Provo Bay, and when they do, they typically subside within two weeks or less [13,34]. This suggests that monitoring and mitigation should be focused on Provo Bay and the marinas for maximum efficacy, and that bloom advisories should be targeted and frequently updated so that recreationists are not dissuaded from using the large areas of the lake that remain unaffected by blooms.

Further research is necessary to provide additional context and support for these findings, especially given the limited scope of remotely-sensed data relative to the number and complexity of environmental drivers of algal blooms. Future work could incorporate machine learning techniques and additional data such as lake levels, precipitation, nutrient loads from the watershed, sediment, and atmospheric deposition, and geochemical and hydrologic modeling of lake conditions. Another important improvement to this work would be the incorporation of imagery from new satellites with higher spectral resolution which can differentiate between green algae and cyanobacterial blooms, since other work suggests that the effects of temperature and turbidity can be different for each species [103,104].

This work demonstrates the general utility of remote sensing data for analyzing spatial and temporal water quality relationships at a lake-wide scale and provides preliminary analyses which can inform Utah Lake management strategies and future research into the lake's unique characteristics and functioning.

Chapter 4: Spatial and Temporal Algal Bloom Dynamics on Utah Lake

This chapter is based on the article “Analysis of spatial and temporal dynamics of algal blooms in Utah Lake using six years of imagery from Sentinel 2” which will be submitted for publication in MDPI Hydrology.

4.1 Introduction

As evidenced by many prior studies of algal blooms and the results presented in the above chapters, blooms are highly temporally variable and spatially heterogeneous, especially on large, shallow, slow-circulating lakes like Utah Lake [2,38,40,59,82,103,105-109].

Understanding where and when algal blooms occur is essential for efficient planning and execution of resource-intensive mitigation and monitoring efforts. In this chapter, we use the chl-a estimates from the previously described point sampling data derived from the satellite imagery (Chapter 2, [110]) to investigate algal bloom dynamics: where and when algal blooms occur on the lake and patterns in their duration and intensity. We develop additional imagery-based data to provide greater detail to the analyses. These results provide insight into where and when bloom monitoring and mitigation efforts could be focused for maximum efficacy, and suggests what a typical growing season on Utah Lake looks like in terms of algal blooms.

4.2 Composite image generation and analysis

To provide additional detail for the spatial analyses, which were inherently limited by the point sampling data despite its comprehensive and random coverage, we generated two additional

datasets which used the entire six-year collection of usable Sentinel 2 images of Utah Lake:

1. A visual classification dataset with qualitative descriptions of bloom patterns in images collected during the growing season, and
2. A composite image dataset which consists of composite images we created using different approaches with Google Earth Engine to calculate summary pixel values for every usable image pixel.

Although the focus of this chapter is “bloom” dynamics, it is important to recognize that there is nothing even approaching consensus in the literature about a specific threshold chl-a concentration which defines the occurrence of an “algal bloom” [108,111,112]. Some suggest that, rather than a universal threshold, blooms should be defined based on deviation from normal conditions for a given waterbody [107,112]. The Utah Division of Water Quality, in their framework for all freshwater bodies in Utah (which includes extremely clear, oligotrophic high alpine lakes), specifies chl-a concentrations above ~7.5 µg/L as a threshold criteria (indirectly, through application of a Carlson State Trophic Index formula) for designating a freshwater body as impaired for recreational use [113]. This may not be a particularly useful number for highly productive, eutrophic Utah Lake, where concentrations are essentially always above that threshold during the growing season, so for this analysis we choose the “deviation from normal” framework and discuss bloom thresholds in terms of percentiles from the distribution of chl-a observations obtained using the methods described in Chapter 2.

To provide additional context on chl-a concentrations and the appearance of algal blooms on Utah Lake, Figure 4.1 shows an unprocessed Sentinel 2 image of Utah Lake taken on September 10th, 2022, when there were active blooms in several areas of the lake, alongside an image showing the results of the chl-a model (described in Chapter 2) applied to each water-containing pixel. The dark areas in the middle of the lake correspond to chl-a concentrations below ~20 µg/L, where the water in the plain image appears slightly green due to background primary productivity in

the water column, but there is no distinct bloom. In Goshen Bay, Provo Bay, and along the eastern shoreline, however, there are regions of distinctly higher chl-a concentrations—approaching 150 $\mu\text{g/L}$ in Provo Bay and the plume extending from it, and exceeding 200 $\mu\text{g/L}$ in Goshen Bay. When viewed from water level, such as from a boat, concentrations in this range would appear as visible scum, mats, or clumps of algae in the water.

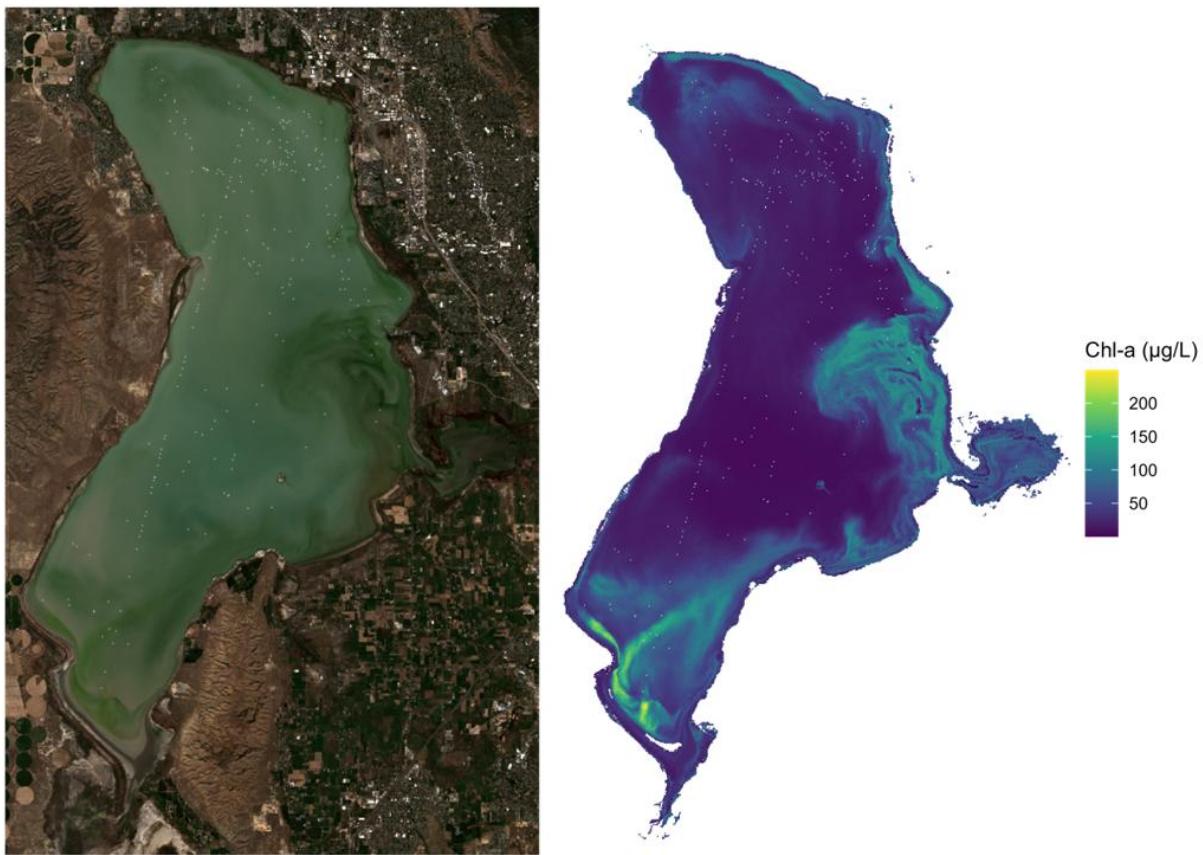


Figure 4.1: Comparison of September 10th, 2022 Sentinel 2 image of Utah Lake with results of chl-a model applied to that image to provide visual context for different levels of chl-a.

4.2.1 Image processing in GEE

For the imagery datasets generated for this analysis, we applied the same image processing steps described in Chapter 2 [13], first developed by Cardall, *et. al.* [11], to the Sentinel 2 image collection,

but instead of extracting data using the sub-sampling approach, we either exported whole images or used the compositing tools available in GEE to create single images that summarize the data for a specific set of images.

For the initial image processing, we applied the necessary band scalers to the Sentinel 2 images, used the Sentinel Scene Classification Layer and the JRC Global Surface Water Mapping layers [66] to mask out areas of land and cloud shadow, and applied the following chl-a retrieval algorithm, developed in Chapter 2, to every usable pixel for every image in the collection:

$$\ln(chla) = 2.90 - 8.95(NDCI) - 17.20(NDCI^2) \\ - 21.96(\rho_{blue}) + 17.11(\rho_{green}). \quad 4.1$$

NDCI is the Normalized Difference Chlorophyll Index developed by Mishra and Mishra [68], which is calculated as

$$NDCI = \frac{\rho_{red} - \rho_{RE}}{\rho_{red} + \rho_{RE}}. \quad 4.2$$

Each image in the Sentinel 2 collection has between 3 and 4 million usable pixels covering Utah Lake, depending on the area of the lake, and with fewer pixels in images with cloud cover. After application of the chl-a retrieval algorithm, the pixel values represented the estimated chl-a concentration at the surface of the water covered by that pixel. The visual classification and compositing techniques applied to the collection of chl-a images are described below.

4.2.1 Visual classification data

The highly variable visual appearance of algal blooms and the frequent presence of cloud cover made it infeasible to analyze the frequency of bloom occurrences in different areas of the lake with traditional quantitative methods. To qualitatively describe bloom genesis, growth, and movement, we visually inspected the chl-a images and recorded the appearance of blooms according to pre-

defined classifications. For this analysis, we used 566 images collected during the growing season (April–October) from 2019–2025. We visualized the images using GEE such that chl-a concentrations above ~30 µg/L were easily visible. We recorded any group of about ~200 adjacent pixels with values at or above this threshold as a bloom, and specified the location of the bloom according to the following classifications:

- Provo Bay Bloom: Algal concentrations above the threshold in any part of Provo Bay
- Provo Bay Plume: Algal concentrations above the threshold following a clear pattern of a plume originating in the Bay and extending from the mouth of Provo Bay into the lake.
- Goshen Bay Bloom: Algal concentrations above the threshold in any part of the southernmost portion of the lake.
- Goshen Bay Plume: Algal concentrations above the threshold extending from the southernmost portion of the lake in a clear plume pattern.
- Eastern Shore Bloom: Algal concentrations above the threshold on the eastern shoreline appearing to be distinct from Provo Bay plume.
- Lake bloom: Algal concentrations above the threshold anywhere else in the lake outside the areas described above.

Figure 4.2 shows representative examples of images from each category. We constructed the dataset with Boolean logic in one-hot columns, so that one image could have a “True” condition for multiple categories. For example, the Provo Bay Plume image (top center) in Figure 4.2 was labelled as “True” for all conditions, as there is a visible bloom in Provo Bay, a bloom extending from Provo Bay into the lake, a bloom in Goshen Bay, a plume extending from Goshen Bay into the lake, and blooms in other areas of the lake distinct from Goshen Bay, Provo Bay, and the Eastern Shoreline. The lower row of images also have blooms in addition to the title bloom. For example, the Goshen Bay Plume image (lower left) also includes a Provo Bay bloom and blooms in the lake. The

Eastern Shore Bloom (lower center) includes a Provo Bay bloom and a Goshen Bay plume. The Other Lake Plume (lower right) includes a Provo Bay Bloom, a Provo Bay Plume, and a Goshen Bay Plume.

Note the small missing areas in the images in Figure 4.2, which were masked out due to clouds or cloud shadow. Not all areas of the lake were visible in every image due to cloud cover and image framing, so we also recorded a variable indicating visibility, and we normalized the results for each category by the number of images in which that location was visible. In addition, some images were completely obscured by clouds or only had a few hundred visible pixels; these were excluded from analysis. Out of 566 growing season images in the Sentinel 2 collection, 402 were included in the final visual classification dataset.

4.2.2 Composite Images

GEE provides a simple process for visualizing summary statistics on a large set of images by generating composite images.

Composite images are created by summarizing the pixel value in each image of a collection by a chosen statistic, such as mean, median, minimum, maximum, or mode or other statistic such as standard deviation. For example, if the chosen statistic is the minimum, the composite image pixel values are the smallest value observed at each pixel over the entire image collection. Figure 4.3 provides a graphical representation of this process, showing how individual images in a collection are summarized by pixel to create a single output image. Composite images are useful for generating a visual representation of spatial differences based on descriptive statistics.

We used the collection of chl-a images to create mean and median composite images for the entire collection, as well as mean images for each month. The total number of images included was 1,238, though individual pixels with clouds, cloud shadow, ice, or land were not included in the calculation, so most pixels had fewer than 1,238 data points. All the pixels, however, had hundreds of values over the six-year study period with which to provide usable statistics.

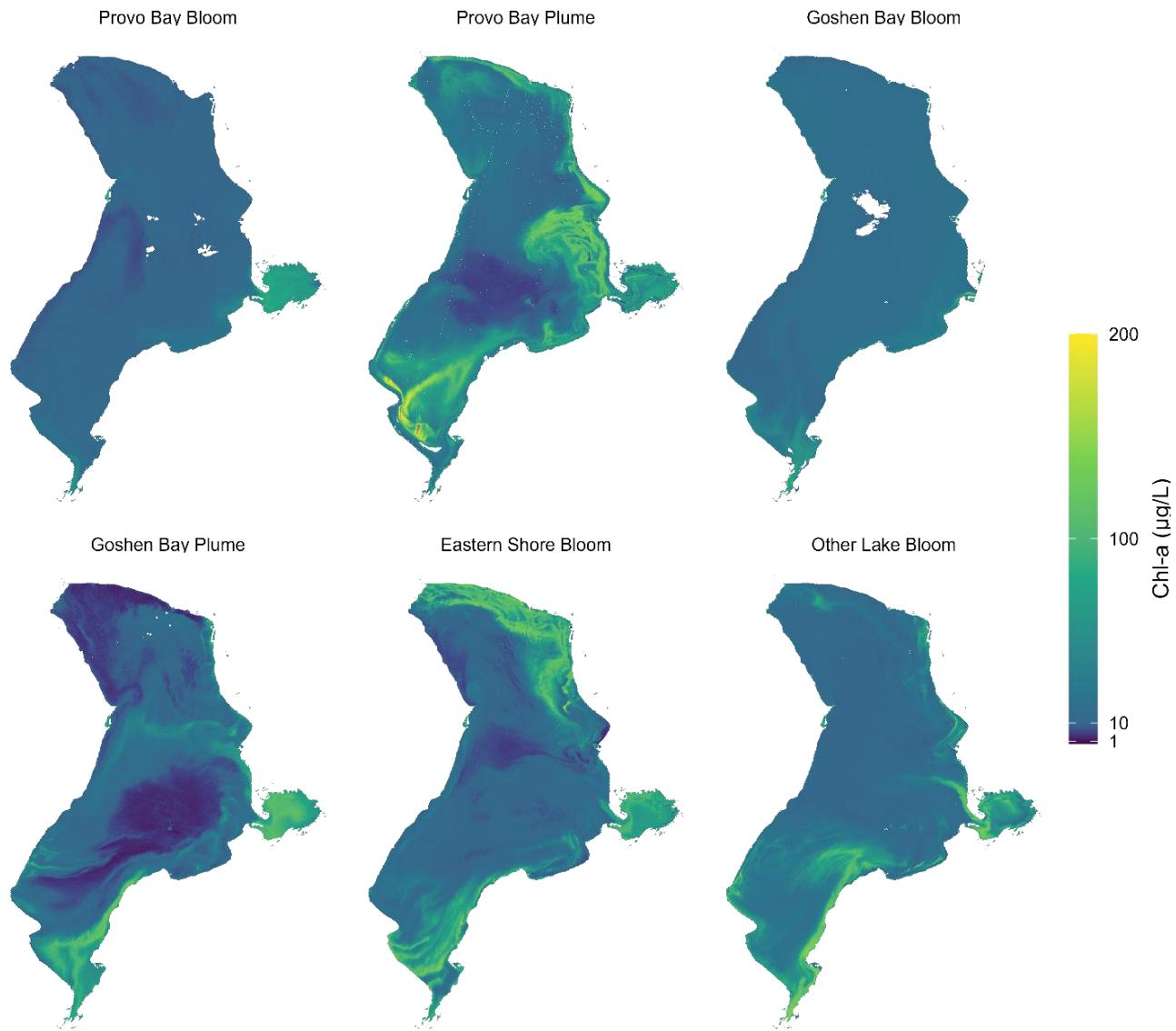
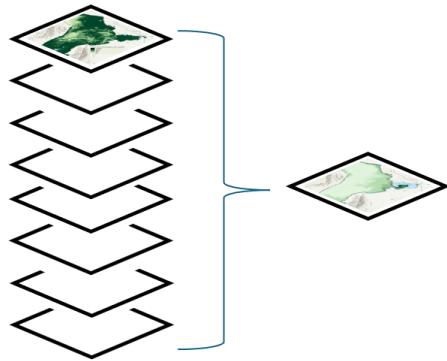


Figure 4.2: Example visual classification images. Note the non-linear color scale to make low-concentration blooms more visible. We used a threshold of $\sim 30 \mu\text{g/L}$ in ~ 200 adjacent pixels to identify blooms.



72

Figure 4.3: Graphical representation of the image compositing process. Individual images are summarized on a per-pixel basis to create a single image.

The mean and median compositing processes generated highly unrealistic chl-a values (in the thousands) for a handful of pixels at the edge of the Sentinel 2 swath, which cuts through Provo Bay. We were unable to identify the cause of this issue, but it may be the result of image overlap between satellite paths. Since it only affected a few of the approximately three million pixels in each image, we simply excluded these unrealistic pixels from analyses.

In addition to the mean and median images, we created a different set of composite images showing the ratio of instances when each pixel was over a specific threshold to the number of observations for that pixel (i.e. the number of images with that pixel visible). We chose the 25th, 50th, and 75th percentile chl-a values of 9 µg/L, 28 µg/L, and 40 µg/L, respectively, as the first three threshold values, and a fourth threshold value of 87 µg/L, which is 1.5 times the IQR above the third quartile, to represent outlier chl-a concentrations. We calculated the values for the percentiles and the outlier threshold using all the non-interpolated chl-a data from the point dataset described in Chapter 2 [110]. We generated the composite images with GEE by summing the number of observations in which each pixel was above the specified value for each threshold and then dividing that sum by the total number of

observations. This method produced images which illustrate spatial differences in how frequently various chl-a concentrations are observed on the lake.

4.3 Composite imagery results

4.3.1 Blooms and chl-a plumes more frequent in shallow bays

We analyzed the visual classification data by creating a table (Table 4.1) of proportions showing the frequency of each phenomenon as a percentage of the images in which the relevant area of the lake was visible. The Lake, Eastern Shore, Goshen Bay, and Provo Bay areas were visible in 402, 349, 289, and 292 images, respectively.

Table 4.1 shows the results of the visual identification of bloom frequency in each area. The total number of images with at least one detectable bloom (i.e. group of ~200 adjacent pixels at chl-a concentrations above ~30 $\mu\text{g/L}$) in a given location was divided by the number of images in which that location was visible to get the frequency.

Table 4.1: Visual image classifications

Area	Bloom Frequency	Plume Frequency
Lake	51% (205 of 402 images)	NA
Eastern Shore	60% (209 of 349 images)	NA
Provo Bay	97% (283 of 292 images)	89% (260 of 292 images)
Goshen Bay	70% (202 of 289 images)	55% (159 of 202 images)

Table 4.1 shows that in about half of the usable images from the growing season, there was a visually-detectable bloom (chl-a concentrations above 30 $\mu\text{g/L}$ in approximately 200 adjacent pixels)

in an area of the lake other than Provo Bay, Goshen Bay, and the eastern shoreline. For the eastern shoreline (both above and below Provo Bay), the frequency rose to slightly above half the time—60%. Images of Provo Bay during the growing season showed chl-a concentrations that met the criteria for a bloom nearly all the time (97%). Goshen Bay was lower than Provo Bay but still had frequent blooms, with 70% of usable images showing a bloom. Similarly, 89% of the usable images where the mouth of Provo Bay was visible showed a distinct plume of chl-a concentrations flowing from the bay out into the lake. For Goshen Bay, the plume was less frequent but still appeared slightly more than half the time in images with the bay visible.

4.3.2 Composite Images

Figure 4.4 shows the median (Panel A) and mean (Panel B) pixel values over the entire image collection.

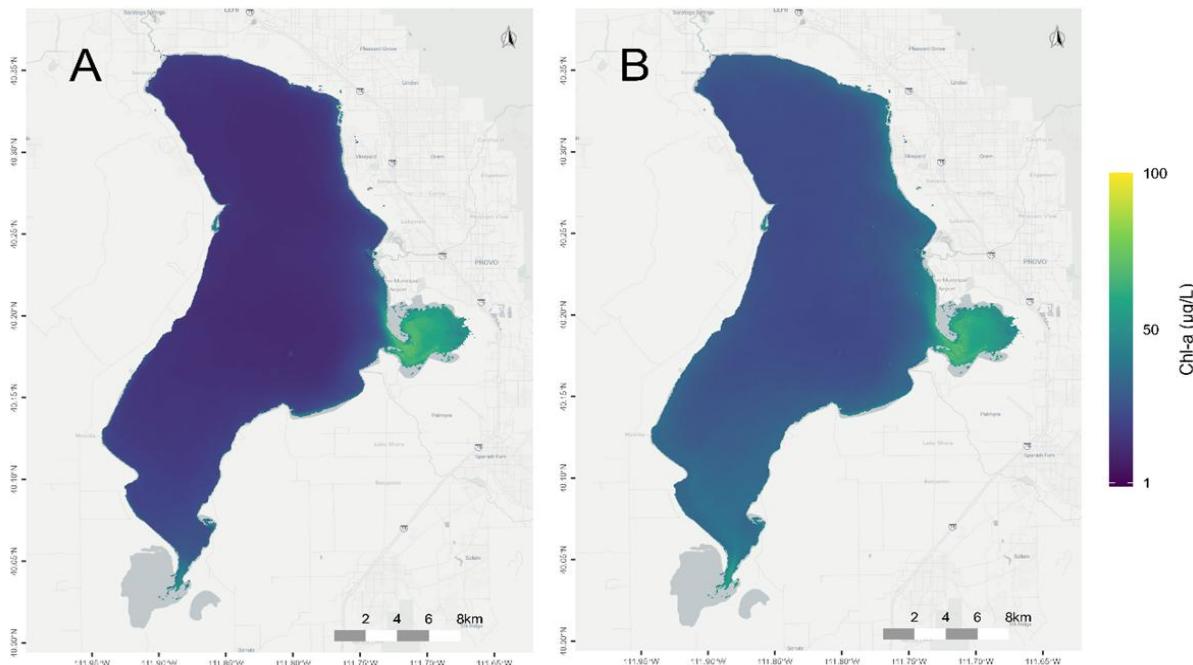


Figure 4.4: Median (A) and mean (B) chl-composite images

From the median and mean images (Panel A and Panel B, respectively), it is immediately obvious that chl-a values for the entire lake are right-skewed, as the mean value is noticeably higher than the median. In the main lake area, the mean lake value is typically double the median value or greater. This suggests that chl-a values on the main lake are typically low (below 25 µg/L) with occasional high values. In contrast, the mean value is typically only about one third higher than the median in Goshen Bay, and only a few µg/L higher in Provo Bay, indicating more frequent high values in these areas which raise the median closer to the mean.

The median value being so much closer to the mean in Provo Bay suggests the chl-a values in Provo Bay are more normally distributed than chl-a values elsewhere on the lake, and remain high throughout the growing season. Both the median and mean images also show chl-a concentrations are typically higher along the eastern shoreline relative to the rest of the lake. All three of these observations align well with the findings from the visual image classification.

Figure 4.5 shows the mean pixel values by growing season months (April-October) over the entire study period. In this case, rather than calculating the mean value for each pixel from every image in the collection, the mean for each month was calculated using only pixel values from that month in every year.

This illustrates prevailing intraseasonal differences in the distribution of chl-a concentrations. The results are mostly unsurprising—chl-a concentrations are low in the spring and peak in August and September before beginning to abate as temperatures drop in the fall. This also clearly shows the phenomenon of Provo Bay developing high chl-a concentrations much earlier in the season than the rest of the lake—with monthly mean values already well over 50 µg/L in April.

Figure 4.6 shows the mean pixel values categorized by year rather than month, illustrating inter-seasonal differences in chl-a concentrations. Interestingly, inter-year chl-a concentrations show nearly as much variation as intra-year (monthly) concentrations, and it is clear that certain phenomena are much stronger in some years than others.

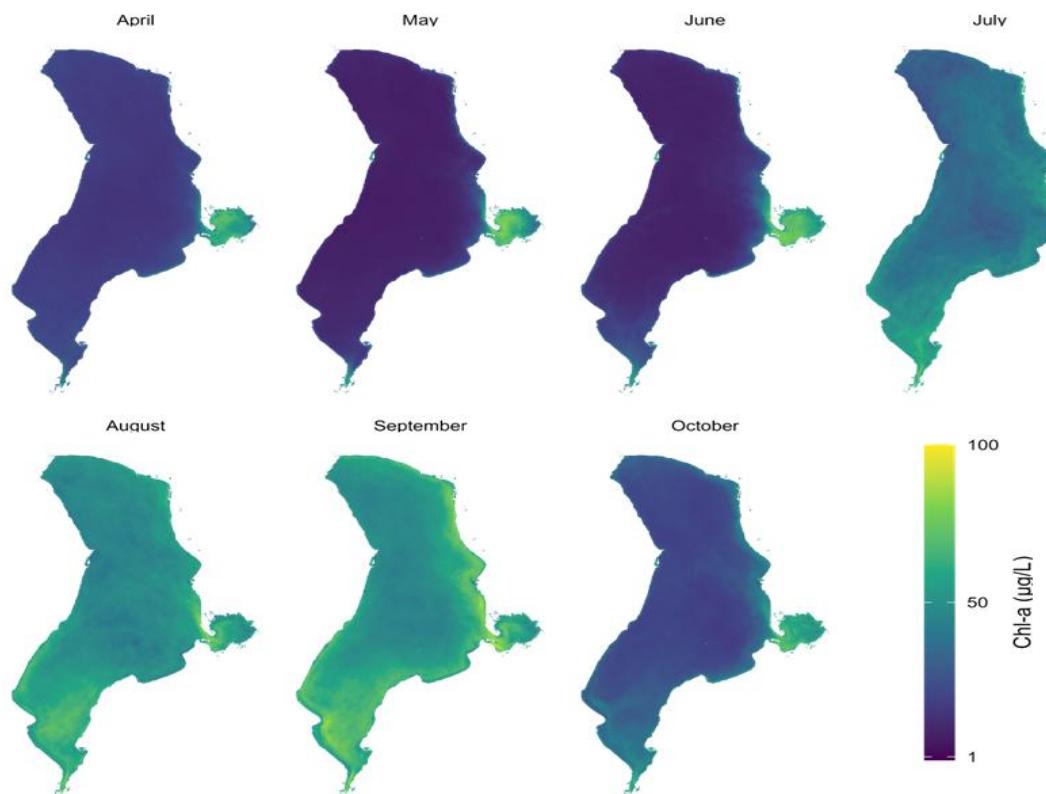


Figure 4.5: Monthly mean composite images

2019 had the lowest chl-a concentrations overall, with mean concentrations ranging from 10 to 12 µg/L in the main body of the lake and 55 to 70 µg/L in Provo Bay. 2021 had the highest concentrations, with mean values ranging from 40 to 50 µg/L in the main body of the lake and 60 to 75 µg/L in Provo Bay, although typical concentrations in Goshen Bay were highest in 2022. The plumes from Provo and Goshen bays and the hotspot along the eastern shore were most intense in 2021 and 2022, though they are still detectible in the images for the other years. This also confirms observations that the main body of the lake (except the area on the east side affected by the Provo Bay plume) generally has the lowest chl-a concentrations.

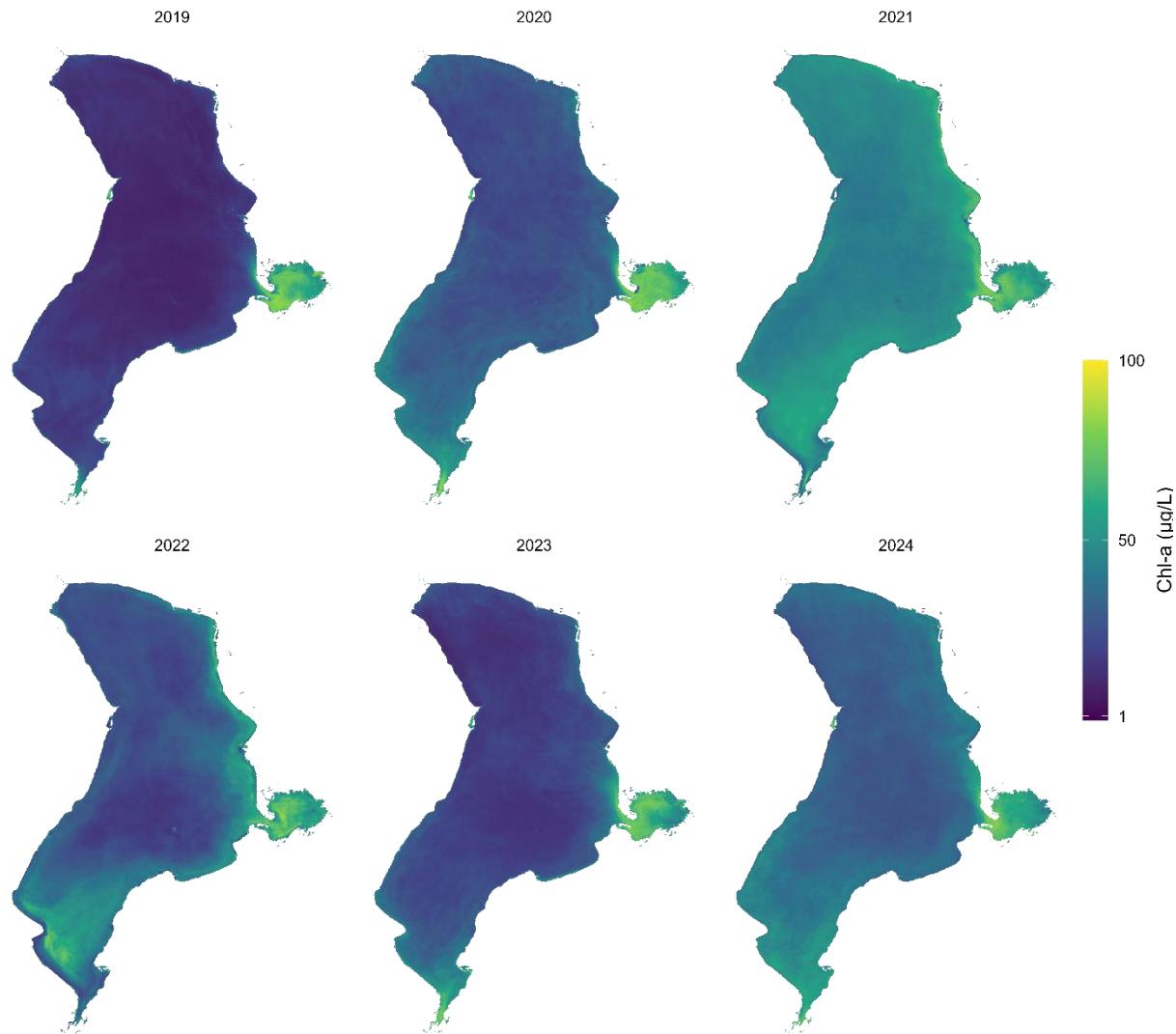


Figure 4.6 Yearly mean chl-a composite images.

Figure 4.7 shows the percentile threshold composite images—i.e. the percentage of observations for each pixel where the pixel value was above the specified percentile. We computed the percentile limits lake-wide using all the point data (Chapter 2).

Matching the observations from the median and mean images, pixels in Provo Bay and Goshen Bay had many more days with chl-a values above the median (50^{th} percentile) value and at outlier levels (1.5 times the IQR above the third quartile). This image also shows that outlier-level chl-a values are almost never observed in the main body of the lake except in the areas immediately outside

Provo Bay and Goshen Bay, which the visual classification data shows are often affected by plumes of high chl-a concentrations extending from these bays.

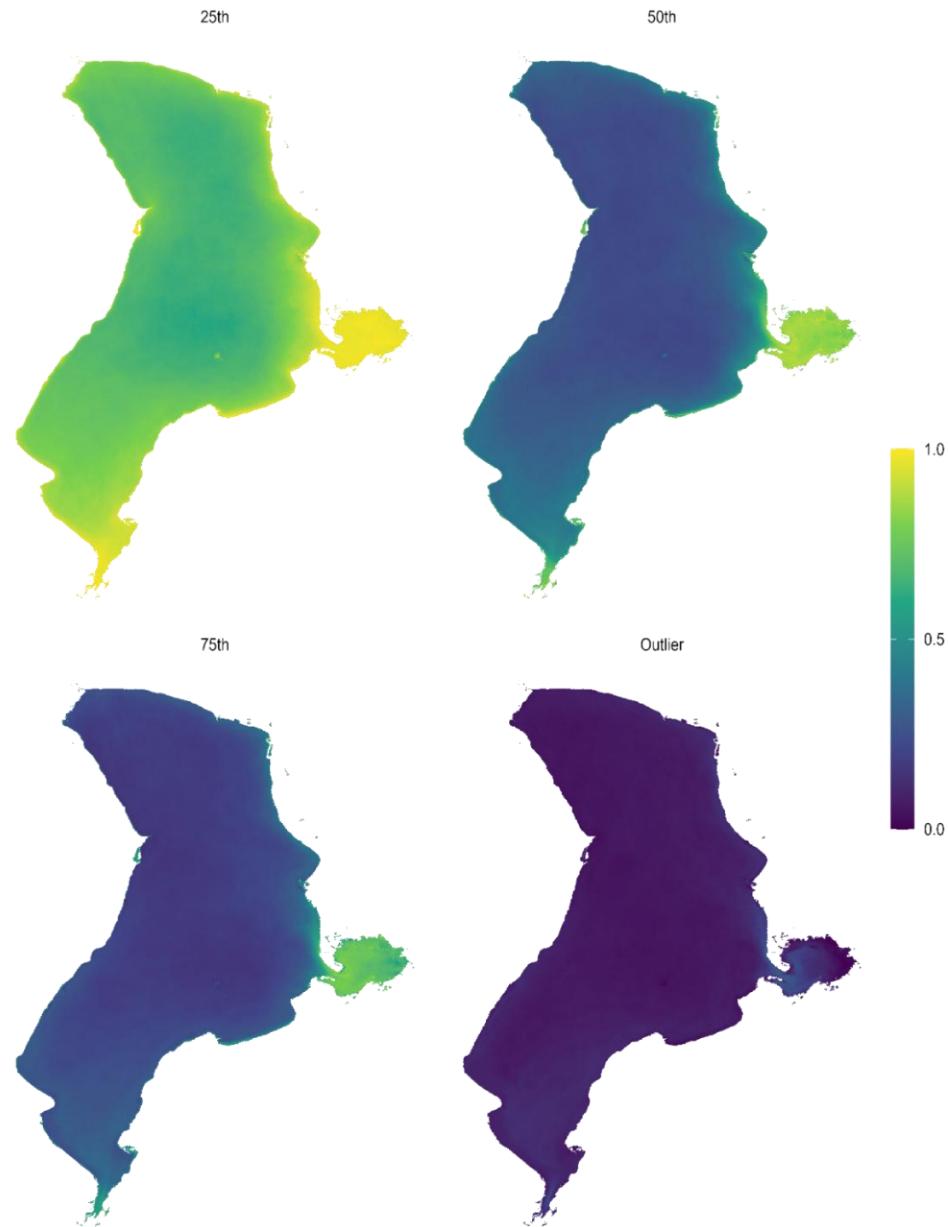


Figure 4.7: Proportion of days above threshold value

The 25th percentile image also shows a distinct buffer of higher percentages that stays relatively close to the shoreline of the lake, with the exceptions of the middle area of the lake affected by the

Provo Bay plume and the southern area affected by the Goshen Bay plume. The central areas of the lake exceeded the 25th percentile value in only about half of the usable pixels, illustrating that blooms in that area are rare relative to the bays and the eastern shoreline area, where concentrations exceeded the median value in almost 75% of the images.

4.4.3 Spatial differences in chl-a value distributions

Figure 4.8 shows histograms of chl-a measurements over the entire study period from the Clusters dataset described in Chapter 2 [64], with Near Shore points located inside Provo Bay visualized as a separate category. The median value for each category is shown as a dotted vertical line, and the mean as a solid line.

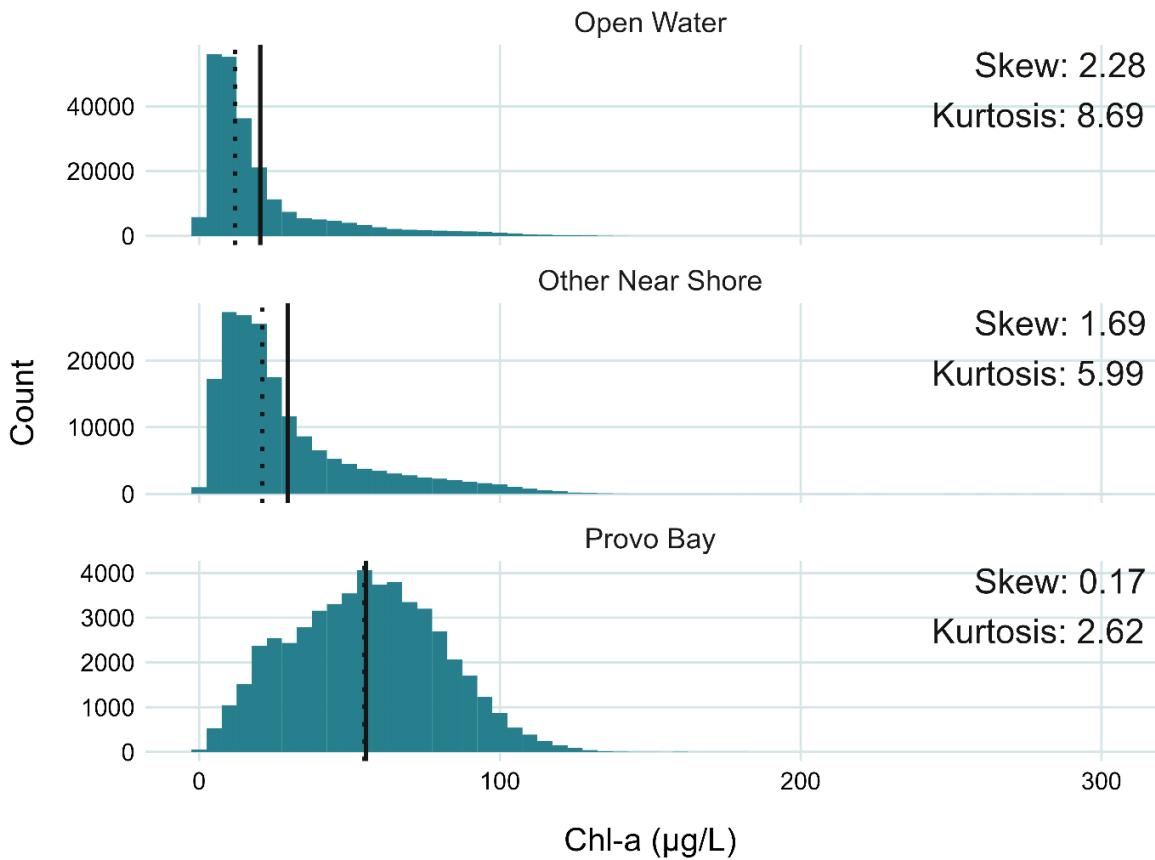


Figure 4.8: Statistical distribution of chl-a measurements by Clusters category. Note: each category has a different number of points (y-axis scales). The “Other Near Shore” category includes all points in the Near Shore category not located inside Provo Bay, and the Provo Bay category includes all points from the Near Shore category that are within the bay (this is different from the Provo Bay category in the Boxes dataset).

Table 4.2 provides the descriptive numbers.

Table 4.2: Distribution descriptions for chl-a from Clusters dataset

	Mean	Median	Skewness	Kurtosis
Open Water	20.33	12	2.28	8.69
Provo Bay	55.47	55	0.17	2.62
Other Near Shore	29.46	21	1.69	5.99

Note that each category has a different number of points (y-axes): there are 100 points in the Open Water category, 77 points in the Other Near Shore category, and 22 points in the Provo Bay category, so the total counts vary accordingly.

Chl-a concentrations in the Open Water category, which represents the main body of the lake (see Section 2.4.3), are extremely right-skewed, with a skewness value of 2.28 and a median value just over half of the mean value. In Near Shore areas excluding Provo Bay, the distribution is still right skewed but less strongly, with a skewness value of 1.69 and a median value that is about two thirds of the mean value. In stark contrast, chl-a concentrations in Provo Bay follow a nearly normal distribution, with a skewness value below one and a mean and median value that are within one $\mu\text{g/L}$ of each other. The mean values for the Open Water, Other Near Shore, and Provo Bay increase from 20.33 to 29.46, to 55.47, respectively with a similar magnitude increase in the median values.

4.5 Discussion

4.5.1 Spatial variation in bloom frequency and intensity

The visual classification analysis shows that, during the growing season (April-October), at least a mild bloom (concentrations at or above $\sim 30 \mu\text{g/L}$) is present somewhere in the open lake about half the time, with more frequent blooms occurring along the eastern shoreline and much more frequently in the shallow Provo and Goshen Bays. It is unclear whether the prevalence of blooms along the eastern shoreline is due to different water column conditions in

that area of the lake, or if algae that begins growing in other areas of the lake accumulates along the eastern shore due to prevailing wind and water circulation patterns.

The high frequency of plumes appearing from Provo Bay (89% of visible images) and Goshen Bay (55% of visible images) suggests that the bays act as “incubators” for algal blooms which then flow out into the main lake. This effect is much stronger in Provo Bay, both because Provo Bay receives the most flow from tributaries, which means more water pushed from inside the bay into the main lake, and because Provo Bay is shallow with clearer water, which promotes algal growth. This suggests that mitigation and monitoring efforts should be focused on the shallow bays and along the eastern shoreline. It also suggests that HAB warnings could be localized to specific areas, since although Provo Bay experiences blooms during almost the entirety of the summer, the open lake is typically free from even mild blooms for at least half of the growing season.

The findings from the visual bloom classification analysis are validated by the percentile frequency composite images, which show that intense blooms on Utah Lake occur predominantly in the shallow bays and near-shore areas. The percentile frequency images showed that chl-a levels over the median of 28 µg/L were observed in the main lake less for less than a third of the year, compared to over half the year in the near shore areas (especially on the eastern shore and southernmost area of the lake) and up to 90% of the year in Provo Bay. Chl-a values above the 75th percentile of 40 µg/L and chl-a values above the outlier level of 87 µg/L were almost exclusively found in Provo Bay and Goshen Bay and immediately outside them.

4.5.2 Temporal patterns

The monthly and yearly mean composite images showed evidence of distinct inter- and intraseasonal patterns and differences in chl-a concentrations on Utah Lake.

The monthly pattern is as expected, with concentrations on most of the lake highest in the late summer and early fall, although average concentrations in Provo Bay are already well over 50 µg/L

in April and remain high through the end of the growing season. Also of note, the average concentration across the entire study period in October is typically higher in most areas on the lake than it was in June, when temperatures are typically warmer. This suggests that colder temperatures in the fall may not affect already-present algal blooms as much as colder temperatures early in the growing season suppress growth.

Also of note, the yearly mean images show nearly as much variation amongst each other as the monthly mean images. There are numerous potential causes for this strong interannual variation, with the two most likely influences being lake levels and growing season temperatures, both of which vary significantly on an annual time scale. For an informal comparison, Table 4.3 provides annual average lake levels and growing season water temperatures obtained from the Utah Division of Water Resources and the MODIS temperature dataset described in Chapter 2, respectively.

Table 4.3: Average annual lake level and growing season water temperature

Year	Average Lake Elevation	Average Growing Season Water Temperature
2019	1366.9 m (4484.7 ft)	20.0°C (68°F)
2020	1367.3 m (4485.9 ft)	22.2 °C (72°F)
2021	1366.7 m (4483.8 ft)	22.2 °C (72°F)
2022	1366.3 m (4482.7 ft)	22.8 °C (73°F)
2023	1366.5 m (4483.3 ft)	21.1 °C (70°F)
2024	1367.8 m (4487.5 ft)	22.2 °C (72°F)

There is no clear correlation between either average water temperature or average lake levels, but low lake levels do seem to be at least somewhat associated with higher prevailing chl-a levels, and low temperatures somewhat associated with lower chl-a levels. 2022 was the lowest water year (lake elevation of 1366.3 m above sea level), and was the year with the most intense blooms in the areas of the lake near the bay, although the lake level in 2021 was similar to the lake level in 2019, and 2021 had the highest chl-a levels in the main lake, while 2019 had the lowest. 2019 and 2023

had the lowest average temperatures, at 20.0 °C and 21.1 °C, respectively, and also appeared to have the lowest average chl-a concentrations overall. However there are distinct overall differences between overall average chl-a concentrations in the 2020 and 2021 images, despite those years having the same average growing season temperature (22.2 °C) and very similar average lake elevations—though the average lake level in 2021 was approximately half a meter (3 feet) lower than in 2020, so there is almost certainly an interactive effect of lake level and average growing season temperature on overall algal growth.

Four of the six complete years in the study period, 2020, 2021, 2022, and 2024, had essentially the same average growing season temperature of 22 °C (72 °F). The composite images for 2020 and 2024 are the most similar of the 6, and also had the two highest lake levels of the six years (1367.3 m and 1367.8 m, respectively). Despite having the same average growing season temperatures as 2021 and 2022, these two years had noticeably lower overall chl-a levels than 2021 and 2022. The difference may be that average lake levels in 2021 and 2022 were the lowest and third-lowest of the six years. There appears to be a fairly strong interactive effect where less algal growth occurs in years with similar average growing season temperatures when lake levels are higher.

Unfortunately, six years of data is not sufficient to quantitatively determine the strength of this pattern, so this is mostly speculation, but it matches existing research that shows blooms thrive in warmer water and shallower water [1,51,94,107].

4.5.3 Extreme value analysis

Analysis of the distributions of chl-a data in different regions of the lake from the points dataset generated in Chapter 2 provided further evidence of the very different behavior of algal blooms in near-shore areas of the lake, especially Provo Bay, relative to areas of open water. The extreme right skew of values from the Open Water category, representing points in the main body of the lake further than ~1km from shore, confirms that blooms on the main lake are quite rare. Data from near-shore areas—within ~1km of

shore, excluding points inside Provo Bay, is less skewed, indicating that blooms are slightly more frequent close to shore. Provo Bay, in contrast, has an approximately normal distribution of chl-a measurements centered around a median of 55 µg/L, indicating that bloom conditions prevail inside the bay for essentially the entirety of the growing season.

4.6 Conclusions

Our analysis of the spatial and temporal dynamics of algal blooms on Utah Lake yielded a mix of expected and surprising results.

As expected, chl-a concentrations peak in the late summer and are generally lowest in the early spring and late fall. Surprisingly, Provo Bay blooms much earlier than the rest of the lake both on a inter- and intraseasonal scale. Chl-a concentrations in Provo Bay rise earlier in the season than they do in the rest of the lake and remain high throughout the summer and late into the fall, such that chl-a concentrations in the bay are normally distributed, rather than right skewed as they are on the rest of the lake. In addition to being a bloom hotspot, blooms often start in Provo Bay and flow out into the open lake, as evidenced by the frequent appearance of a plume from Provo Bay and the fact that the area immediately outside the mouth of the bay had a higher number of days with chl-a levels above the average chl-a value than the rest of the lake.

The eastern shoreline of the lake and Goshen Bay are also bloom hotspots, and Goshen Bay often contributes to blooms on the rest of the lake, as blooms arise inside the bay and then are carried by water circulation and wind into the main body of the lake. This phenomenon is not as prominent in Goshen Bay as it is for Provo Bay, however, as the Provo Bay plume was present in most of the images, while the Goshen Bay plume was only visible in about half.

Although the emphasis of this chapter was on bloom dynamics, we did an informal comparison of the images of the annual chl-a values on the lake with average annual lake level and water temperature to investigate potential relationships between chl-a and lake level and overall growing season temperatures. With only six years of data we are unable to make any conclusive findings,

but it appears chl-a may be sensitive to the combination of lake level and average annual water temperature. The two coldest years, 2019 and 2023, had the lowest average chl-a values lakewide. At the same average temperature, however, the two years with higher lake levels (2020 and 2024) had lower average chl-a concentrations than the two years with lower lake levels (2021 and 2022), suggesting an interactive effect on chl-a from lake level and average water temperature. This could mean that hydrologic management of the lake, which is technically a managed reservoir as there is a dam and pump station at its only outlet, may be a key component of mitigating and managing algal blooms—maintaining higher lake levels could help to prevent excessive algal growth even as global temperatures warm. There are not enough years of data to state this conclusively, however, and further research on longer-running remote sensing datasets would be necessary to confirm this.

Chapter 5: Conclusion

Earth observation satellite data provide an unparalleled opportunity for systems-level analysis of water quality. This is extremely valuable for large freshwater lakes with complex geochemistry and high spatial heterogeneity, like Utah Lake. Remotely-sensed data from earth observation satellites is particularly useful for studying algal blooms thanks to sensors specifically designed for detecting vegetation. The data on algal blooms provided by satellite imagery is much more comprehensive, both spatially and temporally, than the data obtained from in situ samples used in traditional water quality studies, making it extremely valuable for analyzing large-scale, long term trends and behavior. Understanding algal bloom drivers and dynamics for the entire lake is critical for developing effective mitigation and monitoring strategies, and for predicting the success of proposed strategies. This is especially important for Utah Lake, given its ecological and economic significance and its vulnerability to anthropogenic influences due to its proximity to a rapidly growing urban population and its history of pollution and invasive species challenges.

To facilitate greater understanding of algal blooms on Utah Lake, we used high-frequency, high-resolution imagery from Sentinel 2 and MODIS to generate a comprehensive, six-year dataset of chl-a concentrations, water temperature, and turbidity on the lake. We used this dataset to provide key insights into the drivers and spatiotemporal patterns of algal blooms on the entire lake.

We developed a novel approach for retrieving estimates of chl-a concentration and turbidity from Sentinel 2 imagery, combining physics-based equations and regression models built on pairs of in

situ data and satellite observations. We applied these retrieval algorithms to pixel values from three sets of randomly selected points on the lake for each image in the six-year collection of usable images from Sentinel 2 to generate an accessible, comprehensive, and accurate dataset of chl-a and turbidity on Utah Lake. We obtained temperature measurements for the same points from MODIS imagery processed to exclude unrealistic values. We implemented this entire approach in a Jupyter notebook which is publicly available, along with the dataset, to provide scientific transparency and assist other researchers and water managers interested in using remote sensing techniques to study water quality on Utah Lake or other lakes worldwide.

5.1 Key findings

Our analysis of this dataset and additional data generated from the Sentinel 2 imagery yielded several key insights into the behavior of algal blooms on Utah Lake.

Although direct correlations between chl-a concentrations and both water temperature and turbidity are weak and vary in different areas of the lake and at different times of the year, we found that intense blooms (concentrations over 87 µg/L, the outlier threshold for chl-a values in the dataset) were associated with turbidity values below 120 NTU. This finding is confirmed by the observation that Provo Bay experiences earlier, more frequent, and more intense booms despite having typically clearer water, and it suggests that turbidity-related light limitation may constrain the formation of intense algal blooms in Utah Lake despite warm water and high nutrient concentrations that would otherwise accelerate growth.

We also observed a relationship between early-season bloom onset (chl-a concentrations exceeding 34 µg/L) and temperature increases—a disproportionate number of the first blooms occurring during the month of April were preceded by a large temperature jump (relative to the prior 30 days) a few days before the bloom. This suggests that unusually warm days early in the season may act

as a trigger for initial blooms. Blooms later in the season have no observable correlation with temperature changes.

Furthermore, this analysis confirmed and provided details on the pronounced spatial heterogeneity of algal blooms on Utah Lake. We identified Provo Bay as a major bloom hotspot which could be described as an “incubator” for blooms, as there were visually detectable blooms inside the bay in 97% of the images where it was visible during the growing season (April–October), with a plume of high chl-a concentrations extending from the mouth of Provo Bay into the main body of the lake in 89% of the usable images. Goshen Bay is a similar hotspot and incubator, but to a lesser extent, with visible blooms in 70% of the usable images and a plume visible in 55% of the images. In contrast, the main body of the lake had visible blooms in only 51% of the usable images from the growing season—although the eastern shoreline had a slightly higher prevalence of blooms, with a bloom visible somewhere along that shoreline in 60% of the images.

Provo Bay is also distinct from the main lake because chl-a concentrations in the bay follow an approximately normal distribution, unlike the highly right-skewed distribution of chl-a observations on the rest of the lake. The Utah Division of Water Quality considers Provo Bay as a separate water body entirely for the purposes of the Clean Water Act and some management programs [113]. Our findings show this is a valid approach given how differently Provo Bay behaves with respect to algal blooms, but it is also important to recognize that they are not entirely disconnected, especially since one of our most important findings was the observation of the extremely frequent plume of high chl-a concentrations extending from Provo Bay into the main body of the lake.

Another key finding is the predominance of relatively low chl-a concentrations on the main lake outside of Provo Bay and Goshen Bay, especially the areas further than ~1km from shore. Although primary productivity is high and chl-a concentrations are never close to 0 anywhere on the lake, the high concentrations that would be visible as scum, mats, or clumps are very rarely seen on areas of

the open lake, although they are extremely frequent in Provo Bay and Goshen Bay.

A preliminary finding that is not conclusive because a longer study period is necessary to confirm it is that a combination of low lake levels and high temperatures appears to be associated with the most intense growth. Years with similarly high temperatures but higher lake levels had lower overall growth compared to years with the same average growing season temperature but lower lake levels. This suggests higher lake levels may have somewhat of a dampening effect on excessive algal growth that would otherwise occur in response to warm average temperatures

Although cooler annual average temperatures were associated with lower overall chl-a concentrations, and there were some signs that cooler spring temperatures can delay early season growth, there is no strong negative response of algal growth to cold temperatures in the fall. This was demonstrated by both the monthly composite images and the time series of the point sampling dataset. Since the point sampling dataset included the measurements of water temperature, this effect is not simply due to water remaining warmer than the faster-cooling surrounding air—there is likely some other mechanism that allows already-established blooms to remain active even in colder water.

Even when temperatures go back down to the same levels they were in April, chl-a concentrations don't decrease to the levels seen in April, as shown by the time series of the data in Chapter 3 and the monthly composite images in Chapter 4. "This suggests that colder temperatures in the fall may not affect already-present algal blooms as much as colder temperatures early in the growing season suppress growth." The composite imagery does not include temperature information, but the time series data show that water temperatures do decrease while chl-a remains high, indicating that this is not only an effect due to water remaining warm as the weather cools in the fall.

5.2 Error and limitations

As with all studies based on remote sensing data, there is inherent error in our data due to the uncertainties associated with estimating water quality parameters from radiometric data. We applied stringent quality assurance processing on both the Sentinel 2 and MODIS datasets, which were already atmospherically corrected, so the majority of the error for these datasets comes from the chl-a and turbidity retrieval algorithms discussed in Section 2.3. As discussed in that section, the error metrics are good relative to the ranges of both variables, and well below the level at which they would affect any of our conclusions. In addition, we expect most errors to be biased in one direction, which means trends and spatial differences are still represented accurately.

Another technical difficulty with the use of remote sensing data for water quality studies is the fact that the reflected light measured by the satellite sensors is typically only representative of conditions at or near the surface of the water. This means that planktonic algae deeper in the water column and benthic algae will not be detected by satellite sensors, which can lead to underestimation of algal biomass. Although this is a concern and may have biased our data slightly low, conditions on Utah Lake are not favorable for deeper algae and there is likely very little of it. The extremely soft, poorly-defined benthic layer and the feeding activities of carp make it almost impossible for benthic algae to establish, and the prevailing high turbidity means that there is not sufficient light just inches below the water surface for photosynthesis. For these reasons, we assume that, although our method may have missed a small amount of algae, analysis of only surface algae still provides an accurate characterization of algal blooms on Utah Lake.

In a similar vein, another potential issue with detecting water column algae is the fact that submerged aquatic vegetation also contains chl-a, the same pigment we use to indicate algal biomass. For Utah Lake, this is really only an issue in Provo Bay—there are no other areas of the lake with significant established subaquatic vegetation. Provo Bay, however, is heavily vegetated and also less turbid than the main lake, raising the possibility that pixels which read as having high chl-a are actually just pixels with submerged vegetation. We have three reasons for believing that this likely did

not significantly affect our data. First, we performed rigorous quality assurance processing, including excluding data from pixels covering “mixed” areas. Any areas with emergent aquatic vegetation would have been classified as mixed and excluded. Second, as mentioned above, remote sensing data is only representative of conditions at the top of the water surface and not more than a few inches deep into the water column. So unless there was vegetation very close to the surface, it was likely not detected by the satellite. Finally, we closely examined the imagery and data from Provo Bay in order to investigate this issue and assess whether it was affecting our data. We observed that chl-a concentrations were typically high in the central areas of the bay, where there is less vegetation, while areas close to the edge of Provo Bay consistently had lower chl-a concentrations. This is the opposite of what we would expect to see if submerged vegetation were skewing the results, because the vegetation is denser and water is shallower closer to shore. For these reasons, we felt that the Provo Bay data accurately characterizes chl-a concentrations and was not greatly affected by submerged aquatic vegetation.

A non-technical issue but nevertheless a significant limitation on our results is that there are only six years of data from Sentinel 2. This limits our ability to make conclusions about prevailing conditions on the lake, because, as shown in the analyses of inter-seasonal variation, hydrologic and environmental factors can vary greatly between years. The six years included in the dataset fortunately included both very wet and very dry as well as warmer and colder years, and previous research on Utah Lake using 40 years of data from the long-running Landsat mission has shown that average chl-a concentrations on the main lake have remained mostly steady [13]. Taking these things together, we can assume that the six years of data included in this study are reasonably representative of “typical” conditions on Utah Lake, but the strength of our conclusions is nevertheless limited by the fact that this is a relatively small sample of information about the lake on ecological time scales.

Lastly, as with all remote sensing studies, this was a purely observational study. Although we used a random sampling method

for the point dataset, we obviously were not able to isolate or control any of the variables affecting algal growth in Utah Lake, so all of our results are observations of correlations, and cannot be used to prove causation. With complex natural phenomena, however, this is essentially the best we can do, and observations of relationships, patterns, and trends are invaluable for understanding the natural environment.

5.3 Recommendations

Any water quality improvement proposals with the stated goal of increasing the clarity of Utah Lake water should be carefully considered, as the high turbidity of Utah Lake's water column likely inhibits blooms that would otherwise be more severe. It also indirectly suggests that nutrient concentrations are not the only control on algal growth on the lake, meaning that interventions aimed solely at reducing nutrient inputs may not be as effective at controlling HABs on Utah Lake as other methods such as wetland and shoreline restoration.

In addition, although a very preliminary finding based on only a few years of data, the observation that higher lake levels prevent high algal growth that would otherwise occur in response to high average growing season temperatures suggests that hydrologic management may be a critical piece of the Utah Lake management puzzle. Utah Lake is a managed reservoir, and although water rights around Utah Lake are some of the most complex in the world, allocating more resources towards keeping Utah Lake levels high may be essential for reducing HABs on the lake.

The characterization of algal blooms on Utah Lake provided by this study could also help improve the efficiency of monitoring efforts, as it identifies the most likely locations and times for blooms. Based on our findings, for maximum efficacy monitoring efforts should focus on the regions identified as bloom hotspots, specifically Provo Bay, Goshen Bay, and the eastern shoreline. We did not analyze marinas because they were too small to distinguish from shoreline with the resolution of the Sentinel 2 imagery, but based on personal observations during field work, the marinas

behave similarly to the bays, and should be monitored as bloom hotspots. Monitoring campaigns could also be responsive to the 5-day temperature increase that often precedes early season blooms, and could begin monitoring Provo Bay earlier than the rest of the lake because blooms appear in the bay much earlier in the season relative to the rest of the lake.

Importantly, however, sampling campaigns limited to or focused on near-shore areas (including marinas), Provo Bay, and eastern shoreline should not be considered as representative of the lake as a whole, as they will be heavily biased towards high chl-a observations. While focusing monitoring on high chl-a areas may be the most efficient method for issuing timely HAB advisories, researchers and the public should be aware that these datasets and the corresponding public health warnings are not necessarily indicative of the overall health of the lake.

Because our results suggest that intense blooms on the main lake rarely persist for more than a few days, it may be beneficial to issue HAB abatement advisories with the same timeliness and publicity as HAB warnings, and to provide clear information about the location and spatial extent of the HAB. When a HAB advisory is issued because of a bloom along the eastern shoreline or in Provo Bay, people may choose to avoid the lake altogether, even though conditions are perfectly safe just a mile away from shore or at beaches further away from the bloom. More timely, specific guidance about areas of the lake that *are* safe for recreation could improve recreational use and public sentiment about the lake.

Although remote sensing methods involve some error and are unable to resolve causal mechanisms for algal blooms, this research provides useful characterizations of algal bloom dynamics and of the relationship between algal growth and water temperature and turbidity on Utah Lake. Our research methods can serve as a template for future remote sensing studies on freshwater systems besides Utah Lake, and our findings expand scientific understanding of this important natural resource.

References

1. Paerl, H.W.; Otten, T.G. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology* **2013**, *65*, 995-1010, doi:10.1007/s00248-012-0159-y.
2. Anderson, D.M.; Glibert, P.M.; Burkholder, J.M. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries* **2002**, *25*, 704-726, doi:10.1007/bf02804901.
3. Paerl, H.W.; Huisman, J. CLIMATE: Blooms Like It Hot. *Science* **2008**, *320*, 57-58, doi:10.1126/science.1155398.
4. Paerl, H.W.; Huisman, J. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports* **2009**, *1*, 27-37, doi:10.1111/j.1758-2229.2008.00004.x.
5. Abbott, B.W.; Errigo, I.M.; Follett, A.; Lawson, G.; Meyer, M.M.; Moon, H.; Shurtliff, K.; LeMonte, J.J.; Proteau, M.; Davis, K. Getting to know the Utah Lake Ecosystem. *Provo, Utah* **2021**, *32*.
6. Horns, D. Utah Lake comprehensive management plan resource document. *Department of Earth Science, Utah Valley State College, Orem, Utah* **2005**.
7. Williams, G.P. *Great Salt Lake and Utah Lake Statistical Analysis: Vol II: Utah Lake*; Farmington Bay & Utah Lake Water Quality Council: Provo, Ut: 2020.
8. PSOMAS. *Utah Lake TMDL: Pollutant Loading Assessment & Designated Beneficial Use Impairment Assessment*; Department of Environmental Quality 2007.
9. Abu-Hmeidan, H.Y.; Williams, G.P.; Miller, A.W. Characterizing total phosphorus in current and geologic utah lake sediments: Implications for water quality management issues. *Hydrology* **2018**, *5*, 8.
10. Barrus, S.M.; Williams, G.P.; Miller, A.W.; Borup, M.B.; Merritt, L.B.; Richards, D.C.; Miller, T.G. Nutrient Atmospheric Deposition on Utah Lake: A Comparison of Sampling and Analytical Methods. *Hydrology* **2021**, *8*, 123.
11. Cardall, A.; Tanner, K.B.; Williams, G.P. Google Earth Engine Tools for Long-Term Spatiotemporal Monitoring of Chlorophyll-a Concentrations. *Open Water Journal* **2021**, *7*, 4.

12. Olsen, J.; Williams, G.; Miller, A.; Merritt, L. Measuring and Calculating Current Atmospheric Phosphorous and Nitrogen Loadings to Utah Lake Using Field Samples and Geostatistical Analysis. *Hydrology* **2018**, *5*, 45, doi:10.3390/hydrology5030045.
13. Tanner, K.B.; Cardall, A.C.; Williams, G.P. A Spatial Long-Term Trend Analysis of Estimated Chlorophyll-a Concentrations in Utah Lake Using Earth Observation Data. *Remote Sensing* **2022**, *14*, 3664.
14. Cardall, A.C.; Hales, R.C.; Tanner, K.B.; Williams, G.P.; Markert, K.N. LASSO (L1) Regularization for Development of Sparse Remote-Sensing Models with Applications in Optically Complex Waters Using GEE Tools. *Remote Sensing* **2023**, *15*, doi:10.3390/rs15061670.
15. Brahney, J. Estimating total and bioavailable nutrient loading to Utah Lake from the atmosphere. **2019**, 1-31, doi:<https://doi.org/10.15142/em0e-tt24>.
16. Brimhall, W.H., Merritt, Lavere B. Geology of Utah Lake: Implications for Resource Management. *Great Basin Naturalist Memoirs* **1981**, *5*.
17. Hogsett, M.; Li, H.; Goel, R. The Role of Internal Nutrient Cycling in a Freshwater Shallow Alkaline Lake. *Environmental Engineering Science* **2018**, *36*, 551-563, doi:10.1089/ees.2018.0422.
18. Heckmann, R.A.; Thompson, C.W.; White, D.A. Fishes of Utah Lake. *Great Basin Naturalist Memoirs* **1981**, 107-127.
19. Merritt, L.B. Utah Lake: A Few Considerations. **2017**.
20. Lawson, G.; Daniels, J.; Jones, E.F.; Buck, R.; Baker, M.; Abbott, B.; Aanderud, Z. Utah Lake's Cyanobacteria Proliferation and Toxin Production in Response to Nitrogen and Phosphorous Additions. **2020**.
21. Lawson, G.M. Seasonal Nutrient Limitations of Cyanobacteria, Phytoplankton, and Cyanotoxins in Utah Lake. Brigham Young University, 2021.
22. Liljenquist, G.K. *Study of Water Quality of Utah Lake Tributaries and the Jordan River Outlet for the Calibration of the Utah Lake Water Salinity Model (LKSIM)*; Brigham Young University: 2012.
23. Miller, S.A., Crowl, T.A. Effects of common carp (*Cyprinus carpio*) on macrophytes and invertebrate communities in a shallow lake. *Freshwater Biology* **2006**, *51*, 85-94, doi:<https://doi.org/10.1111/j.1365-2427.2005.01477.x>.
24. Barnes, A.J. *The effect of the Goshen Bay dike on the benthos of Utah Lake in relation to water quality*; Brigham Young University: 1974.
25. Richards, D. *Chlorophyll A trends in Utah Lake from 1989 to 2019*; 2022.
26. Richards, D.C. Plankton Biomass, Diets, Production-Biomass Ratios, and Ecotrophic Efficiency Estimates for Utah Lake Foodweb Model Development. **2021**, doi:<https://doi.org/10.13140/RG.2.2.14446.54080>.
27. Richards, D.C. *Seasonal patterns of phytoplankton assemblage densities and functional traits in Utah Lake*; OreoHelix Ecological: Vineyard, UT, 2021.

28. Richards, D.; Miller, T. *Ecological health and integrity of Utah Lake Progress Report 2019* version 2.3; Vineyard, UT, 2019.
29. Richards, D.; Miller, T. *Utah Lake Progress Report 2017-2018: Chapter 1 Phytoplankton Assemblages*; Vineyard, UT, 2019.
30. Richards, D. *Development of Primary Production-Light Limitation Metrics for Monitoring Water Quality in Utah Lake*; Vineyard, UT, 2021.
31. Rushforth, S.R.; Squires, L.E. New records and comprehensive list of the algal taxa of Utah Lake, Utah, USA. *The Great Basin Naturalist* **1985**, 237-254.
32. Shiozawa, D.K.; Barnes, J.R. The Microdistribution and Population Trends of Larval Tanypus Stellatus Coquillett and Chironomus Frommeri Atchley and Martin (Diptera: Chironomidae) in Utah Lake, Utah. *Ecology* **1977**, 58, 610-618, doi:<https://doi.org/10.2307/1939010>.
33. Strong, A.E. Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah Lake. *Remote Sensing of Environment* **1974**, 3, 99-107, doi:[https://doi.org/10.1016/0034-4257\(74\)90052-2](https://doi.org/10.1016/0034-4257(74)90052-2).
34. Tate, R.S. *Landsat collections reveal long-term algal bloom hot spots of Utah Lake*; Brigham Young University: 2019.
35. Williams, R.R.R. Determining the anthropogenic effects on eutrophication of Utah Lake since European settlement using multiple geochemical approaches. Brigham Young University, 2021.
36. Zanazzi, A.; Wang, W.; Peterson, H.; Emerman, S.H. Using Stable Isotopes to Determine the Water Balance of Utah Lake (Utah, USA). *Hydrology* **2020**, 7, 88.
37. Lawson, G.M.; Young, J.L.; Aanderud, Z.T.; Jones, E.F.; Bratsman, S.; Daniels, J.; Malmfeldt, M.P.; Baker, M.A.; Abbott, B.W.; Daly, S.; et al. Nutrient limitation and seasonality associated with phytoplankton communities and cyanotoxin production in a large, hypereutrophic lake. *Harmful Algae* **2025**, 143, 102809, doi:<https://doi.org/10.1016/j.hal.2025.102809>.
38. Han, Y.; Aziz, T.N.; Del Giudice, D.; Hall, N.S.; Obenour, D.R. Exploring nutrient and light limitation of algal production in a shallow turbid reservoir. *Environmental Pollution* **2021**, 269, 116210, doi:<https://doi.org/10.1016/j.envpol.2020.116210>.
39. Llames, M.E.; Lagomarsino, L.; Diovisalvi, N.; Fermani, P.; Torremorell, A.M.; Perez, G.; Unrein, F.; Bustingorry, J.; Escaray, R.; Ferraro, M.; et al. The effects of light availability in shallow, turbid waters: a mesocosm study. *Journal of Plankton Research* **2009**, 31, 1517-1529, doi:10.1093/plankt/fbp086.
40. Liu, X.; Chen, L.; Zhang, G.; Zhang, J.; Wu, Y.; Ju, H. Spatiotemporal dynamics of succession and growth limitation of phytoplankton for nutrients and light in a large shallow lake. *Water Research* **2021**, 194, 116910, doi:<https://doi.org/10.1016/j.watres.2021.116910>.
41. Brown, R. Relationships Between Suspended Solids, Turbidity, Light Attenuation, and Algal Productivity. *Lake and Reservoir Management* **1984**, 1, 198-205, doi:10.1080/07438148409354510.

42. Williams, R.; Nelson, S.; Rushforth, S.; Rey, K.; Carling, G.; Bickmore, B.; Heathcote, A.; Miller, T.; Meyers, L. Human-Driven Trophic Changes in a Large, Shallow Urban Lake: Changes in Utah Lake, Utah from Pre-European Settlement to the Present. *Water, Air, & Soil Pollution* **2023**, 234, 218, doi:10.1007/s11270-023-06228-5.
43. Wood, R. Acute animal and human poisonings from cyanotoxin exposure — A review of the literature. *Environment International* **2016**, 91, 276-282, doi:<https://doi.org/10.1016/j.envint.2016.02.026>.
44. Gaddis, E.B.; Phillips-Barnes, J. Harmful Algal Bloom Program 2021 Update. **2021**.
45. Winslow, B. Environmental experiments on Utah Lake could lead to solutions to restore it. *FOX 13 News Utah (KSTU)* **2023**.
46. Merritt, L.B.; Miller, A.W. *Interim Report on Nutrient Loadings to Utah Lake: 2016*; Jordan River, Farmington Bay & Utah Lake Water Quality Council: Provo, UT, October 2016.
47. Taggart, J.B.; Ryan, R.L.; Williams, G.P.; Miller, A.W.; Valek, R.A.; Tanner, K.B.; Cardall, A.C. Historical Phosphorus Mass and Concentrations in Utah Lake: A Case Study with Implications for Nutrient Load Management in a Sorption-Dominated Shallow Lake. *Water* **2024**, 16, doi:10.3390/w16070933.
48. Huisman, J.; Weissing, F.J. Light-Limited Growth and Competition for Light in Well-Mixed Aquatic Environments: An Elementary Model. *Ecology* **1994**, 75, 507-520, doi:<https://doi.org/10.2307/1939554>.
49. Urabe, J.; Sterner, R.W. Regulation of herbivore growth by the balance of light and nutrients. *Proceedings of the National Academy of Sciences* **1996**, 93, 8465-8469, doi:10.1073/pnas.93.16.8465.
50. Torremorell, A.; Bustigorry, J.; Escaray, R.; Zagarese, H.E. Seasonal dynamics of a large, shallow lake, laguna Chascomús: The role of light limitation and other physical variables. *Limnologica* **2007**, 37, 100-108, doi:<https://doi.org/10.1016/j.limno.2006.09.002>.
51. Burford, M.A.; Hamilton, D.P.; Wood, S.A. Emerging HAB Research Issues in Freshwater Environments. In *Global Ecology and Oceanography of Harmful Algal Blooms*, Gilbert, P.M., Berdalet, E., Burford, M.A., Pitcher, G.C., Zhou, M., Eds.; Springer International Publishing: Cham, 2018; pp. 381-402.
52. Brezonik, P.; Menken, K.D.; Bauer, M. Landsat-based Remote Sensing of Lake Water Quality Characteristics, Including Chlorophyll and Colored Dissolved Organic Matter (CDOM). *Lake and Reservoir Management* **2005**, 21, 373-382, doi:10.1080/07438140509354442.
53. Brivio, P.A.; Giardino, C.; Zilioli, E. Validation of satellite data for quality assurance in lake monitoring applications. *Science of the total environment* **2001**, 268, 3-18.
54. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R.V. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. *Water* **2020**, 12, doi:10.3390/w12010169.
55. Abirhire, O. Phytoplankton dynamics in relation to turbidity and other environmental factors in Lake Diefenbaker. **2023**.

56. Yuan, X.Y.; Wang, S.R.; Fan, F.Q.; Dong, Y.; Li, Y.; Lin, W.; Zhou, C.Y. Spatiotemporal dynamics and anthropologically dominated drivers of chlorophyll-a, TN and TP concentrations in the Pearl River Estuary based on retrieval algorithm and random forest regression. *Environmental Research* **2022**, *215*, doi:10.1016/j.envres.2022.114380.
57. Shi, K.; Zhang, Y.; Qin, B.; Zhou, B. Remote sensing of cyanobacterial blooms in inland waters: present knowledge and future challenges. *Science Bulletin* **2019**, *64*, 1540-1556, doi:10.1016/j.scib.2019.07.002.
58. Yip, H.D.; Johansson, J.; Hudson, J.J. A 29-year assessment of the water clarity and chlorophyll-a concentration of a large reservoir: Investigating spatial and temporal changes using Landsat imagery. *Journal of Great Lakes Research* **2015**, *41*, 34-44, doi:<https://doi.org/10.1016/j.jglr.2014.11.022>.
59. Tate, R.S. Landsat Collections Reveal Long-Term Algal Bloom Hot Spots of Utah Lake. **2019**.
60. Hansen, C. *Google Earth Engine as a Platform for Making Remote Sensing of Water Resources a Reality for Monitoring Inland Waters*; 2015.
61. Hansen, C.; Swain, N.; Munson, K.; Adjei, Z.; Williams, G.P.; Miller, W. Development of Sub-Seasonal Remote Sensing Chlorophyll-A Detection Models. *American Journal of Plant Sciences* **2013**, *04*, 21-26, doi:10.4236/ajps.2013.412a003.
62. Hansen, C.H.; Burian, S.J.; Dennison, P.E.; Williams, G.P. Evaluating historical trends and influences of meteorological and seasonal climate conditions on lake chlorophyll a using remote sensing. *Lake and Reservoir Management* **2019**, *36*, 45-63, doi:10.1080/10402381.2019.1632397.
63. Hansen, C.H.; Williams, G.P. Evaluating Remote Sensing Model Specification Methods for Estimating Water Quality in Optically Diverse Lakes throughout the Growing Season. *Hydrology* **2018**, *5*, doi:10.3390/hydrology5040062.
64. Tanner, K.B.; Cardall, A.C.; Williams, G.P. A Six-Year, Spatiotemporally Comprehensive Dataset and Data Retrieval Tool for Analyzing Chlorophyll-a, Turbidity, and Temperature in Utah Lake Using Sentinel and MODIS Imagery. *Data* **2025**, *10*, doi:10.3390/data10080128.
65. Hansen, C.H. Google Earth Engine as a Platform for Making Remote Sensing of Water Resources a Reality for Monitoring Inland Waters. In Proceedings of the World Environmental and Water Resources, Austin, TX, May 17-25, 2015.
66. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418-422, doi:10.1038/nature20584.
67. Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* **2016**, *16*, doi:10.3390/s16081298.

68. Mishra, S.; Mishra, D.R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment* **2012**, *117*, 394-406, doi:<https://doi.org/10.1016/j.rse.2011.10.016>.
69. Hansen, C.H.; Williams, G.P. Evaluating Remote Sensing Model Specification Methods for Estimating Water Quality in Optically Diverse Lakes throughout the Growing Season. *Hydrology* **2018**, *5*, doi:10.3390/hydrology5040062.
70. PSOMAS, S. Utah Lake TMDL: Pollutant Loading Assessment & Designated Beneficial Use Impairment Assessment. *Prepared for the Utah Division of Water Quality. Salt Lake City, UT* **2007**, 1-88.
71. Fabian Pedregosa, G.V., Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825--2830.
72. Cruz-Retana, A.; Becerril-Piña, R.; Fonseca, C.R.; Gómez-Albores, M.A.; Gaytán-Aguilar, S.; Hernández-Téllez, M.; Mastachi-Loza, C.A. Assessment of Regression Models for Surface Water Quality Modeling via Remote Sensing of a Water Body in the Mexican Highlands. *Water* **2023**, *15*, doi:10.3390/w15213828.
73. Nahorniak, J.S.; Abbott, M.R.; Letelier, R.M.; Scott Pegau, W. Analysis of a Method to Estimate Chlorophyll-a Concentration from Irradiance Measurements at Varying Depths. *Journal of Atmospheric and Oceanic Technology* **2001**, *18*, 2063-2073, doi:[https://doi.org/10.1175/1520-0426\(2001\)018<2063:AOAMTE>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<2063:AOAMTE>2.0.CO;2).
74. Hansen, C.H.; Williams, G.P.; Adjei, Z.; Barlow, A.; Nelson, E.J.; Miller, A.W. Reservoir water quality monitoring using remote sensing with seasonal models: case study of five central-Utah reservoirs. *Lake and Reservoir Management* **2015**, *31*, 225-240, doi:10.1080/10402381.2015.1065937.
75. Lazhu; Yang, K.; Qin, J.; Hou, J.; Lei, Y.; Wang, J.; Huang, A.; Chen, Y.; Ding, B.; Li, X. A Strict Validation of MODIS Lake Surface Water Temperature on the Tibetan Plateau. *Remote Sensing* **2022**, *14*, doi:10.3390/rs14215454.
76. Pour, H.K.; Duguay, C.R.; Solberg, R.; Rudjord, Ø. Impact of satellite-based lake surface observations on the initial state of HIRLAM. Part I: evaluation of remotely-sensed lake surface water temperature observations. *Tellus A: Dynamic Meteorology and Oceanography* **2014**, *66*, 21534.
77. Tavares, M.H.; Cunha, A.H.F.; Motta-Marques, D.; Ruhoff, A.L.; Cavalcanti, J.R.; Fragoso Jr, C.R.; Martín Bravo, J.; Munar, A.M.; Fan, F.M.; Rodrigues, L.H.R. Comparison of methods to estimate lake-surface-water temperature using Landsat 7 ETM+ and MODIS imagery: Case study of a large shallow subtropical lake in southern Brazil. *Water* **2019**, *11*, 168.

78. Chavula, G.; Brezonik, P.; Thenkabail, P.; Johnson, T.; Bauer, M. Estimating the surface temperature of Lake Malawi using AVHRR and MODIS satellite imagery. *Physics and Chemistry of the Earth, Parts A/B/C* **2009**, *34*, 749-754.
79. Hook, S.J.; Vaughan, R.G.; Tonooka, H.; Schladow, S.G. Absolute radiometric in-flight validation of mid infrared and thermal infrared data from ASTER and MODIS on the Terra spacecraft using the Lake Tahoe, CA/NV, USA, automated validation site. *IEEE Transactions on Geoscience and Remote Sensing* **2007**, *45*, 1798-1807.
80. Pareeth, S.; Salmaso, N.; Adrian, R.; Neteler, M. Homogenised daily lake surface water temperature data generated from multiple satellite sensors: A long-term case study of a large sub-Alpine lake. *Scientific Reports* **2016**, *6*, 31251.
81. Crosman, E.T.; Horel, J.D. MODIS-derived surface temperature of the Great Salt Lake. *Remote Sensing of Environment* **2009**, *113*, 73-81.
82. Liu, G.; Ou, W.; Zhang, Y.; Wu, T.; Zhu, G.; Shi, K.; Qin, B. Validating and mapping surface water temperatures in Lake Taihu: Results from MODIS land surface temperature products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2015**, *8*, 1230-1244.
83. Pelleg, D.; Moore, A. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *Machine Learning*, p **2002**.
84. Virtanen, P.a.G., Ralf and Oliphant, Travis E. and; Haberland, M.a.R., Tyler and Cournapeau, David and; Burovski, E.a.P., Pearu and Weckesser, Warren and; Bright, J.a.v.d.W., St{\e}fan J. and; Brett, M.a.W., Joshua and Millman, K. Jarrod and; Mayorov, N.a.N., Andrew R. J. and Jones, Eric and; Kern, R.a.L., Eric and Carey, C J and; Polat, I.l.a.F., Yu and Moore, Eric W. and; {VanderPlas}, J.a.L., Denis and Perktold, Josef and; Cimrman, R.a.H., Ian and Quintero, E. A. and; et al. {SciPy} 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261--272, doi:10.1038/s41592-019-0686-2.
85. Zhou, Z.-X.; Yu, R.-C.; Zhou, M.-J. Resolving the complex relationship between harmful algal blooms and environmental factors in the coastal waters adjacent to the Changjiang River estuary. *Harmful Algae* **2017**, *62*, 60-72, doi:<https://doi.org/10.1016/j.hal.2016.12.006>.
86. Meyer, B.S.; Heritage, A.C. Effect of Turbidity and Depth of Immersion on Apparent Photosynthesis in Ceratophyllum Demersum. *Ecology* **1941**, *22*, 17-22, doi:10.2307/1930005.
87. Jewson, D.H.; Taylor, J.A. The influence of turbidity on net phytoplankton photosynthesis in some Irish lakes. *Freshwater Biology* **1978**, *8*, 573-584, doi:<https://doi.org/10.1111/j.1365-2427.1978.tb01479.x>.
88. Waker, M.J.; Robarts, R.D. Microbial nutrient limitation in prairie saline lakes with high sulfate concentration. *Limnology and Oceanography* **1995**, *40*, 566-574, doi:<https://doi.org/10.4319/lo.1995.40.3.0566>.

89. Dokulil, M.T. Environmental control of phytoplankton productivity in turbulent turbid systems. In *Phytoplankton in Turbid Environments: Rivers and Shallow Lakes: Proceedings of the 9th Workshop of the International Association of Phytoplankton Taxonomy and Ecology (IAP) held in Mont Rigi (Belgium), 10–18 July 1993*, Descy, J.-P., Reynolds, C.S., Padisák, J., Eds.; Springer Netherlands: Dordrecht, 1994; pp. 65-72.
90. Ho, J.C.; Michalak, A.M. Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes. *Limnology and Oceanography* **2020**, *65*, 992-1009, doi:<https://doi.org/10.1002/lno.11365>.
91. Zhao, D.; Huang, J.; Li, Z.; Yu, G.; Shen, H. Dynamic monitoring and analysis of chlorophyll-a concentrations in global lakes using Sentinel-2 images in Google Earth Engine. *Science of The Total Environment* **2024**, *912*, 169152, doi:<https://doi.org/10.1016/j.scitotenv.2023.169152>.
92. Abbott, B.W. Seven problems with the Utah Lake islands proposal. *approximately limitless* **2021**, 2024.
93. Zou, W.; Xu, H.; Zhu, G.; Zhu, M.; Guo, C.; Xiao, M.; Zhang, Y.; Qin, B. Why do algal blooms intensify under reduced nitrogen and fluctuating phosphorus conditions: The underappreciated role of non-algal light attenuation. *Limnology and Oceanography* **2023**, *68*, 2274-2287, doi:<https://doi.org/10.1002/lno.12421>.
94. Paerl, H.W.; Huisman, J. Blooms Like It Hot. *Science* **2008**, *320*, 57-58, doi:10.1126/science.1155398.
95. Kim, K.B.; Jung, M.-K.; Tsang, Y.F.; Kwon, H.-H. Stochastic modeling of chlorophyll-a for probabilistic assessment and monitoring of algae blooms in the Lower Nakdong River, South Korea. *Journal of Hazardous Materials* **2020**, *400*, 123066, doi:<https://doi.org/10.1016/j.jhazmat.2020.123066>.
96. Pinheiro, J.a.B., Douglas and DebRoy, Saikat and Sarkar,; Team}, D.a.R.C. {nlme}: Linear and Nonlinear Mixed Effects Models. **2021**.
97. Zeileis, A.; Hothorn, T. Diagnostic Checking in Regression Relationships. *R News* **2002**, *2*, 7-10.
98. Lewis, F.; Butler, A.; Gilbert, L. A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution* **2011**, *2*, 155-162, doi:<https://doi.org/10.1111/j.2041-210X.2010.00063.x>.
99. Team, R.C. R: A Language and Environment for Statistical Computing. **2023**.
100. Taggart, J.B.; Ryan, R.L.; Williams, G.P.; Miller, A.W.; Valek, R.A.; Tanner, K.B.; Cardall, A.C. Historical Phosphorus Mass and Concentrations in Utah Lake: A Case Study with Implications for Nutrient Load Management in a Sorption-Dominated Shallow Lake. *Water* **2024**, *16*, 933.
101. Yip, H.; Johansson, J.; Hudson, J. A 29-year assessment of the water clarity and chlorophyll-a concentration of a large reservoir: Investigating spatial and temporal changes using Landsat imagery. *Journal of Great Lakes Research* **2015**, *41*, 34-44.

102. Department of Environmental Quality, U. *Utah Lake Water Quality Study: Phase 1 Report*; Utah, USA, 2018.
103. Bertani, I.; Steger, C.E.; Obenour, D.R.; Fahnensiel, G.L.; Bridgeman, T.B.; Johengen, T.H.; Sayers, M.J.; Shuchman, R.A.; Scavia, D. Tracking cyanobacteria blooms: Do different monitoring approaches tell the same story? *Science of The Total Environment* **2017**, *575*, 294-308, doi:10.1016/j.scitotenv.2016.10.023.
104. Calderini, M.L.; Pääkkönen, S.; Salmi, P.; Peltomaa, E.; Taipale, S.J. Temperature, phosphorus and species composition will all influence phytoplankton production and content of polyunsaturated fatty acids. *Journal of Plankton Research* **2023**, *45*, 625-635, doi:10.1093/plankt/fbad026.
105. Jia, T.; Zhang, X.; Dong, R. Long-Term Spatial and Temporal Monitoring of Cyanobacteria Blooms Using MODIS on Google Earth Engine: A Case Study in Taihu Lake. *Remote Sensing* **2019**, *11*, 2269, doi:10.3390/rs11192269.
106. Le, C.; Li, Y.; Zha, Y.; Wang, Q.; Zhang, H.; Yin, B. Remote sensing of phycocyanin pigment in highly turbid inland waters in Lake Taihu, China. *International Journal of Remote Sensing* **2011**, *32*, 8253-8269, doi:10.1080/01431161.2010.533210.
107. Carstensen, J.; Henriksen, P.; Heiskanen, A.-S. Summer algal blooms in shallow estuaries: Definition, mechanisms, and link to eutrophication. *Limnology and Oceanography* **2007**, *52*, 370-384, doi:<https://doi.org/10.4319/lo.2007.52.1.0370>.
108. Ho, J.C.; Michalak, A.M. Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie. *Journal of Great Lakes Research* **2015**, *41*, 317-325, doi:10.1016/j.jglr.2015.01.001.
109. Klemas, V. Remote sensing of algal blooms: an overview with case studies. *Journal of Coastal Research* **2012**, *28*, 34-43.
110. Tanner, K.B., Cardall, A.C., Williams, G.P. A six-year, spatiotemporally comprehensive dataset and data retrieval tool for analyzing chlorophyll-a, turbidity, and temperature in Utah Lake using Sentinel and MODIS imagery. *Data* **2025**.
111. Oliver, R.L.; Ganf, G.G. Freshwater Blooms. In *The Ecology of Cyanobacteria: Their Diversity in Time and Space*, Whitton, B.A., Potts, M., Eds.; Springer Netherlands: Dordrecht, 2002; pp. 149-194.
112. Gorney, R.M.; Graham, J.L.; Murphy, J.C. The "H," "A," and "B" of a HAB: A definitional framework. *Lakeline* **2023**, *43*, 7-11.
113. Quality, U.D.o.E. *2024 Integrated Report on Water Quality*; Salt Lake City, UT, 2024.

Appendix A: Chl-a and turbidity plots by month

