USING REMOTE SENSING TO EVALUATE HISTORICAL TRENDS AND CONTRIBUTING FACTORS TO ALGAL BLOOM DYNAMICS AND FORECASTING FUTURE CONDITIONS IN THE GREAT SALT LAKE SYSTEM

by

Carly Hyatt Hansen

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Civil and Environmental Engineering

The University of Utah

December 2018

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Carly Hyatt Hansen**

has been approved by the following supervisory committee members:

Steven Burian	, Chair	**9/11/2018** Date Approved
Christine Pomeroy	, Member	**8/23/2018** Date Approved
Michael Barber	, Member	**8/3/2018** Date Approved
Philip Dennison	, Member	**9/7/2018** Date Approved
Gustavious Williams	, Member	**9/7/2018** Date Approved

and by **Michael Barber** , Chair/Dean of

the Department/College/School of **Civil and Environmental Engineering**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Various processes and factors contribute to the occurrence, timing, magnitude, and extent of algal blooms in the Great Salt Lake (GSL) system (including the Great Salt Lake, Farmington Bay, and Utah Lake). While this system is recognized for its role in local and global ecosystems, recreation, and industry, the practices of monitoring, assessing, and planning for changing water quality conditions are severely limited in their ability to describe relationships between the many contributing factors and algal bloom conditions. The aim of this work is to use traditional field sampling and remotely sensed records to explore patterns and trends and identify some of the key factors that influence or contribute to algal bloom conditions in the lakes of the GSL system. Factors explored in this study include local weather, seasonal climate, and hydrologic variables, which have particular relevance to the nearby developing urban area that is experiencing uncertain and changing climate conditions. The study is divided into three distinct bodies, which enables a more complete examination of historical algal bloom patterns, the processes that influence them, and uses this information to guide monitoring and management practices in the future. This research brings together a wide breadth of data types and sources to gain a more holistic view of the complex lake system. The three major objectives of this dissertation are to: 1) evaluate historical patterns and trends using remotely sensed estimates of algal biomass; 2) describe the complex relationships between climate and hydrologic variables and algal blooms through a data-driven modeling and analysis approach; 3) use these relationships to develop a decision support framework that can be used to forecast conditions within the lake system. Primary impacts of this work include an improved understanding of historical water quality conditions, context for evaluating ongoing conditions, knowledge of how external factors contribute to and influence these conditions, and tools for better planning and monitoring practices in the future.

To Adeline for being my motivation and my distraction.

TABLE OF CONTENTS

Appendices

LIST OF TABLES

Tables

ACKNOWLEDGMENTS

I would like to acknowledge and give thanks to the excellent mentors and advisors I have had over the years, particularly: Dr. Gus Williams, who first gave me an opportunity to participate in research and trusted me with keys to the boat; and Dr. Steve Burian for always providing advice, opportunities to challenge myself and learn new skills, and for sharing his passion for teaching others. I also thank other members of my committee who provided input and support throughout my program, especially Dr. Phil Dennison who is an exemplary teacher and provided valuable feedback.

I would like to acknowledge the many individuals who have supported this work through data sharing and providing feedback, particularly those at the USGS Water Science Center in Salt Lake City and Ben Holcomb and others at the Utah Division of Water Quality. I would also like to thank other members of the Urban Water Research Group at the University of Utah for valuable feedback, and Sara Larsen at the Western States Water Council for her mentorship and providing additional opportunities to develop technical skills and broaden my understanding of challenges in water resources.

I also appreciate the encouragement and support of friends and family who motivated me, listened, and encouraged me to overcome the challenges of my program. Finally, I am especially grateful to my husband, Bradley, for his support and confidence in my abilities and his interest in my work. I couldn’t ask for a better editor, field assistant, audience, and partner.

CHAPTER 1

INTRODUCTION

1.1 Harmful Algal Blooms: A Growing Concern for Lakes

Inland lakes provide numerous benefits to the people and ecosystems surrounding them: supplying drinking water, providing opportunities for recreation and habitat for aquatic organisms, fish, and birds, and serving as receiving waters for treated wastewater and stormwater runoff. In recent years, lakes across the United States and around the world have received increased attention and public concern regarding harmful algal blooms (HABs) and eutrophication (Smith 2003, Brooks et al. 2017).

HABs consist of either photosynthetic algae whose cellular structures or collective biomass negatively affect food web dynamics or protozoans with the ability to produce toxins (Anderson et al. 2002). Algae do have an important role in lake food webs; however, not all types of algae can be digested by grazing zooplankton and other aquatic organisms (e.g. some diatoms with hard outer shells, or some species of cyanobacteria) and when algae experience excessive growth (known as “blooms”), this can have a variety of negative impacts on water quality. Algal blooms can cause anoxic conditions when the algae die and are decomposed, depleting the dissolved oxygen that is needed by other aquatic organisms (Marcarelli 2001, Wurtsbaugh 2008). They can also negatively impact the lake ecosystem by limiting light penetration for benthic plants

(Paerl 1988, Paerl et al. 2001, Smith 2003, Paerl and Otten 2013). Limited light penetration reduces biomass and productivity of oxygen-producing benthic plants, which further restricts available dissolved oxygen (sometimes resulting in fish kills) and can limit growth of competing species of algae that are at lower depths in the water column. When competing algae populations and zooplankton are limited, this can result in a positive feedback where growth of the bloom-forming species is further promoted. In addition to these ecosystem impacts, HABs also have consequences for public health and recreation. Some bloom-forming species can produce hepatoxins and neurotoxins at concentrations that result in illness and mortality in small mammals and birds, and illness in humans from inhalation or contact (from rashes to respiratory problems) (Hunter 1998, Backer et al. 2010, Carmichael et al. 2016). Lake aesthetics and recreation can also be impacted through the formation of surface scums, and as anoxic conditions encourage the growth of sulfate reducing bacteria, production of hydrogen sulfide gas, and cause an unpleasant rotten-egg odor (Acuña 2017, Chen et al. 2017). In drinking water sources, HABs can affect taste and odor, requiring additional treatment or use of alternative sources.

Some researchers have suggested that HABs may be one of the greatest water quality threats to public health and aquatic ecosystems in inland waters (Brooks et al. 2016) and have provided evidence that the problem is increasing on a global scale (Hallegraeff 1993). In the Great Lakes region for example, reports of human and mammal illness attributed to HABs have been on the rise (Carmichael 2016). Throughout the United States, issues associated with algal blooms and eutrophication of surface waters are estimated to cost over 2 billion dollars annually (Dodds et al. 2009) due to lost revenue from recreation, costs for treatment, and restoration. In areas where there is likely influence of human activity on HABs (rather than strictly naturally occurring HABs), determining the true cost of algal blooms is highly complex. It includes not only surplus costs for treatment and costs due to lost revenue, but also the costs of how society values the beneficial uses of the lakes given the costs and requirements of various mitigation measures (Hoagland and Scatasta 2006). For the GSL system, there may be differences in how recreation and wastewater dilution are valued in light of management and mitigation alternatives. These management and mitigation alternatives, which can be extremely costly, include urban and watershed management, such as major retrofits to wastewater facilities or managing stormwater inputs to the lake. Other alternatives include physical controls (e.g. aeration, controlled mixing and circulation, or surface skimming), chemical controls (e.g. applying algaecides or coagulation/flocculation), or biological controls (e.g. artificial wetlands and management of species that contribute to mixing and sediment resuspension like bottom-dwelling carp) (Kim 2006).

Despite concerns from research communities, monitoring and regulatory agencies, and the public, the patterns of HAB dynamics (magnitude, frequency, and duration) and the factors that contribute to HABs are not well understood. The United States Harmful Algal Blooms and Hypoxia Research and Control Amendments Act, passed in 2014, highlighted the need for additional understanding of HAB events, their causes, and consequences (US Congress 2014). While this act focuses on the Great Lakes region, the need for improved understanding of HABs is apparent for inland lakes throughout the United States (Brooks et al. 2016).

1.2 HABs in the Great Salt Lake System

HABs are a major water quality concern in the complex system of lakes in central Utah, especially the Great Salt Lake, Farmington Bay, and Utah Lake (referred to hereafter as the Great Salt Lake (GSL) system). Historically, this lake system has not benefited from the extensive efforts and resources of large-scale programs like those in the Great Lakes or other regions. While the lakes in the GSL system do not provide drinking water, there are major concerns over the recreational, aesthetic, public health, and ecosystem and habitat impacts of HABs. This section provides a brief overview of the GSL system, its unique characteristics, and history of HABs.

1.2.1 Description of the Study Area

The GSL system is a remnant of the ancient fresh-water Lake Bonneville, which once stretched across Idaho, Utah, and Nevada. The present-day GSL is located in north-central Utah, between the Wasatch and Oquirrh Mountains. As Lake Bonneville drained over the past 16,000-17,000 years, it left the GSL system, which includes the Great Salt Lake, the Jordan River, and Utah Lake to the south (Arnow and Stephens 1990), which are shown in Figure 1.1.

The lake system has changed a great deal since early settlement of Salt Lake Valley in the late 1840s-1850s; many of these changes have been engineered, including construction of railroad and automobile causeways or operation of systems of pumps, drains, and diversions. Other changes to the system have resulted from climate variability and intense urbanization from Salt Lake City to Provo. Both natural processes (e.g. precipitation, evaporation, snowmelt, and streamflow) and anthropogenic practices (e.g.

water diversions, wastewater and stormwater discharge, and infrastructure operation) contribute to the hydrology and characteristics of the lake water quality (Arnow 1984, Arnow and Stephens 1990).

The GSL is often considered as two distinct lakes, separated by the east-west railroad causeway, forming a Northern Arm (Gunnison, Willard, and Bear River Bays) and a less saline Southern Arm (Gilbert and Farmington Bays), as illustrated in Figure 1.1. Gilbert and Farmington Bays are the terminal points for several major rivers: Jordan, Weber, and Ogden. Also shown in Figure 1.1 are the Jordan River and Utah Lake, key components of the GSL system that connect the major contributing streams of the Spanish Fork and Provo Rivers to the southern end of Farmington Bay. Utah Lake and the Jordan River are both heavily managed, with the Jordan River flows being pumped and regulated by a number of canal and irrigation companies and water conservancy districts (Hooton 1989).

The lakes that form the GSL system have major social, economic, and environmental importance on local, regional, and even global scales. Locally, the recreation from several state parks draws thousands of visitors to these lakes each year. Each of the lakes support a vibrant and diverse ecosystem for millions of migratory birds (Cox and Kadlec 1995) and feature a number of popular recreation and camping spots. Additionally, the large number of land and water-access areas make these lakes popular hunting and bird-watching destinations. Utah Lake is an important year-round habitat for waterfowl and a number of fish, including the endangered June Sucker (Andersen 2007). Additionally, the GSL system provides a significant net economic value to publicly owned treatment works by serving as a receiving water body for wastewater discharge.

For the GSL alone, this benefit is estimated to be as much as $58.9 million annually. The total contribution of all services and uses of the GSL (industrial, aquaculture, and recreational) to the Utah Gross Domestic Product is estimated at $1.3 billion annually (Bioeconomics Inc. 2012). Major connections between algae, human activity, watershed interactions, beneficial uses of the lakes, the aquatic food web, and climate in this system are represented in Figure 1.2. The variety of interactions between the urban area and the lake system means that there are a wide variety of stakeholders with competing interests and views on how to deal with HABs. This can result in contention between stakeholders, where there is disagreement or lack of information about the history and current state of water quality in these lakes, which often stymies discussion about mitigation or future (potentially costly) water management strategies. Thus, the value, benefits derived from these surface waters by the developing urban area, and the potentially massive costs for mitigation motivates further exploration of HABs in these lakes.

1.2.2 Previous Study of HABs in the GSL System

Historical monitoring and sampling efforts from state and federal agencies have helped broadly define algae populations in this system. One of the most common species of algae in Farmington Bay are a toxic cyanobacteria, *Nodularia* (Wurtsbaugh 2008, Wurtsbaugh et al. 2012) whereas the most prevalent algae in the GSL are diatoms and green algae. Historically, diatoms have dominated Utah Lake, though many species of chlorophyta (green algae) and cyanobacteria (blue-green algae) have also been documented (Rushforth and Squires 1985). Sampling during recent large blooms in Utah Lake revealed extremely high cyanobacteria cell counts and elevated levels of toxins at several locations (Utah DEQ 2016). Cyanobacteria blooms in this system and other nearby lakes have been shown to cause illnesses and rashes on swimmers in several Utah lakes, and they have even been the cause of several canine deaths (Wurtsbaugh 2008, Penrod 2015). Both toxic and nontoxic HABs have impacted the GSL system by negatively affecting lake aesthetics (Nicholson and Marcarelli 2004, GSLEP 2018).

The direct drivers of algae growth (in general and in the GSL system) have been widely explored through field and laboratory analysis. Algae growth is primarily a function of available nutrients, light, and temperature. Characteristics of the algae (e.g. speciation, diversity, spatial density, and distribution within the water column) can also be influenced by other factors such as salinity and pH. Generally, it is recognized that both climate and hydrologic processes (urban and natural) influence these factors and affect the growth of algae in receiving water bodies (Price 2011); however, the degree of influence is uncertain (Anderson 2005), and the understanding of external factors, especially in this system, is limited. In the GSL system, studies over the past decade have primarily focused on the relationships between algae and internal factors that contribute to the patterns of speciation and proliferation of algae biomass. These studies have suggested connections between algal speciation and salinity, nutrients, water temperature, etc. (Wurtsbaugh and Berry 1990, Goel and Myers 2009, Wurtsbaugh et al. 2012, Larson and Belovsky 2013, Marden 2015, Wurtsbaugh 2015, Merritt 2017). While one recent study of Utah Lake did explore external factors, it was severely limited by exploring these factors in the context of a single bloom event (Page et al. 2018). A review of major studies and significant findings related to algal bloom characteristics, patterns, and the influencing internal factors is provided in Table 1.1.

This study demonstrates an approach for using remote sensing and statistical models to evaluate past, present, and future conditions of HABs, enabling a more holistic understanding of water quality. While the demonstration focuses on the particular region of the GSL system, the methods are broadly applicable and could be used in other areas with limited monitoring/historical records and areas facing threats of increasing HABs. Beyond the GSL system, researchers have recognized the need to better understand the past to better anticipate and predict conditions in the future (Heisler et al. 2008, Anderson 2009, Anderson et al. 2012). Again, while the information about trends and behavior is specific to the GSL system, the methodology is broadly applicable to any inland surface water region with limited historical and current monitoring records.

Similar to the focus of field studies in the GSL system, predictive modeling of HABs has been largely concentrated on internal factors and the biological and physical processes of algae growth using process-based models or statistical models. This approach requires extensive knowledge of other water quality constituents, hydrodynamics, and characteristics of the algae (speciation, growth and grazing rates, etc.). Examples of these types of internally-focused models include estuarine and lake HAB-related indices (Nojavanet et al. 2014, Obenour et al. 2014, Forio et al. 2015) and algae speciation and biomass in both lakes (Malve et al. 2007) and estuaries (Alameddine et al. 2011). Adding to the current understanding of HAB dynamics and the factors that influence them requires alternative methods for representing HABs and exploring other types of influencing factors, such as more indirect external hydrologic and climate conditions. These other types of factors are particularly relevant to lake water quality and HABs given projected changes in climate and hydrologic conditions and increasing urbanization of the watersheds surrounding this lake system (Naftz et al. 2008, Whitehead et al. 2009).

1.3 Research Objectives, Goals, and Hypotheses

The following section describes specific research objectives and the hypotheses that will be tested by the proposed research activities outlined in Section 1.3. The objectives are divided into three major sections: 1) Enhancing the Historical Record of HABs via Remote Sensing, 2) Improving Understanding of Contributing Factors to HABs, and 3) Developing a Data-driven Forecasting Framework to Support Proactive Monitoring Strategies.

1.3.1 Enhancing the Historical Record of HABs via Remote Sensing

Over the past decade, remote sensing models have been used in large-scale applications that provide valuable spatial and temporal information that supplement existing field sampling data. In the GSL system, the existing history of HABs consists of field samples from several organizations that are irregular, incomplete, and provide limited information about the variability and trends of water quality conditions within the system and even within the lakes. Remote sensing is suggested as a means of extending the limited GSL system water quality record and overcoming the spatial limitations of traditional monitoring techniques. The first objective of this dissertation is to use satellite remote sensing to extend the spatial and temporal record of HABs (as measured via chlorophyll *a* (chl- *a*)) in the GSL system. This objective involves 1) determining a technique for remote sensing that is appropriate for the study area and available data, 2) developing and applying a remote sensing model to historical data to produce an extended record, and 3) using the remote sensing models to explore spatial and temporal behavior in the remotely sensed water quality estimates. These tasks address the recommendations for increased exploration of spatial patterns of HABs in general (Michelakaki and Kitsiou 2005, Wang and Liu 2005) and specifically within the GSL system (Wurtsbaugh 2015). The enhanced record will enable trend analysis and exploration of which scales these trends occur. The hypothesis of this work is that *the lakes in the Great Salt Lake system demonstrate long-term eutrophication through an increase in chl-a concentrations, spatial extent, and variability in the historical record.*

1.3.2 Improving Understanding of Contributing Factors to HABs

A number of recent studies have explored the links between specific factors such as nutrient loading, salinity, or temperature and field-measured or laboratory-simulated conditions to algae populations in the lakes of the GSL system (see Table 1.1). However, these studies are limited in the factors considered, and the scale at which the relationships between HABs and influencing factors are explored. Therefore, this next objective is to use the enhanced spatial and temporal record provided by remote sensing to further explore factors that have historically influenced the algal blooms in the GSL system.

This objective entails 1) identifying the short-term influence of local climate on HAB dynamics and 2) reviewing patterns and relationships between these climate factors and algal bloom dynamics over the historical record. In the Great Lakes region, a variety of factors have been identified as contributing to algal bloom severity and occurrence, including: intense storm events, lack of rainfall, local wind and temperatures, surface runoff, nonpoint source pollution, and anthropogenic loading of nutrients (Michalak et al. 2013, Steffen et al. 2014, Ho and Michalak 2017). Currently, the level of influence of these factors on conditions in the GSL area is highly uncertain. The influence of these potentially contributing factors, the variety of stakeholders, and costs associated with mitigation strategies underscores the need to better understand potential factors for the specific area and their influence on HABs.

Anecdotal evidence in the study area supports the idea that the climate factors have had varying levels of influence throughout the GSL system. For example, scums of algae have been observed to be transported from contained bays in Utah Lake to open water; however, it is unclear how blooms in other parts of the lake are affected by the same conditions. There is also a good deal of uncertainty surrounding the relative impact of external factors such as climate conditions or hydrologic inputs compared to anthropogenic activities. The hypothesis for this objective is that *HAB measures in Utah Lake are more sensitive to local climate and meteorological conditions than in other lakes of the GSL system.* This work is an important extension of the literature by exploring additional measures of HABs, factors that immediately affect these HAB measures, and those that have connections to HAB measures over a long period of time.

1.3.3 Developing a Data-driven Forecasting Framework to Support Proactive Monitoring Strategies

The final portion of the dissertation addresses the need to shift monitoring and management strategies towards anticipating and planning for future conditions. This paradigm shift places an emphasis on preparation, rather than response to poor water quality conditions. To address this shifting perspective, the final objective of this dissertation is to *forecast the likelihood of future bloom events within the study area*.

This objective is accomplished through a forecasting framework *that demonstrates how seasonal predictions of the occurrence of HABs can be used to guide monitoring and management plans*. In water resources, decision support tools are increasingly being used to identify areas for targeted mitigation or to evaluate outcomes and consequences of different management strategies (Obropta et al. 2008, Fischer et al. 2017). Rather than evaluating tradeoffs between competing strategies of various stakeholders, the forecasting framework suggested in this dissertation focuses on physical aspects (hydrological, meteorological, and water quality) and social aspects (acceptable risk of a HAB event) to inform and support monitoring decisions and strategies.

This work builds on previous sections of the dissertation, using data and relationships derived from other research objectives to create a forecasting model. The forecasting model uses information from important external factors that are regularly monitored. The monitoring-support framework uses web services to couple the predictive model with observational data from third-party databases to produce forecasts of future likelihood of an event. The end result is demonstrated in a simple, operational hydroinformatic support tool that can be used to guide monitoring decisions.

1.4 Dissertation Outline

1.4.1 Chapter 2

This chapter lays a foundation for developing remote sensing techniques for evaluating HABs in the GSL system. A unique approach to field sampling is demonstrated using frequent, repeated samples throughout the growing season and spatial offsets, to evaluate the spatiotemporal patterns within this unique lake system. The observed lake characteristics and patterns provide a basis for many of the assumptions used in empirical, lake-specific model development and applications discussed in subsequent chapters. Despite the increasing use and number of applications of remote sensing of water quality, there are significant concerns associated with remote sensing model development and application, several of which are addressed in this chapter. These concerns include: (1) selecting sensors that are suitable for the spatial and temporal variability in the water body; (2) determining appropriate uses of near-coincident data in empirical model calibration; and (3) recognizing potential limitations of remote sensing measurements that are biased toward surface and near-surface conditions. These issues are addressed through sampling at scales that are representative of commonly used sensors, repeated sampling, and sampling at both near-surface depths and throughout the water column. The variability across distances is representative of the spatial resolutions of common sensors (i.e. Landsat, SENTINEL-2, and MODIS) and justify the use of these sensors in the GSL lake system. The observed temporal variability highlighted potential challenges in the system, with temporal patterns proving to be complex and highly variable conditions in some portions of the lake system, making it difficult to detect short events. Temporal variability patterns in these lakes are also used to assess near-coincident data in empirical model development, which is integral to the model development approach used in Chapter 3. Finally, this chapter includes an evaluation of relationships between surface and water column conditions, which helps illustrate potential issues with near-surface remote sensing, particularly when there are events that cause mixing in the water column.

1.4.2 Chapter 3

Building on the results of the study in Chapter 2, this chapter demonstrates a methodology for remote sensing that is grounded in observations of local, lake-specific processes and patterns of variability. Historical satellite imagery, which is becoming increasingly easy to access and use, provides a means for estimating historic conditions or filling gaps in currently collected data. This chapter demonstrates the use of satellite imagery (specifically from Landsat sensors) to create an historic record of algal biomass (using chl- *a* as a proxy measurement) and presents a workflow for developing seasonal algal estimation models using open source tools. This workflow utilizes the Google Earth Engine platform for data collection and R statistical software for processing and model development. This approach is both physically-based (using observed patterns of variability and algal succession in the lake) and data-driven (relying on statistical methods for model development). It produces lake-specific, seasonal models for each of the complex water bodies in the multilake system of the Great Salt Lake and Utah Lake. The 32-year record produced by this methodology is then used to identify long-term trends and relationships to local climate conditions and events. This chapter illustrates a number of ways the long-term trends and patterns can be used to inform monitoring efforts as well as regulatory and development policies.

1.4.3 Chapter 4

This chapter further explores the influence of climate and hydrologic variables on algal blooms and how these various factors can be used to predict and plan for algal blooms, specifically for Utah Lake. Historical relationships between influencing factors and remotely sensed chl- *a* concentrations (and the lake's corresponding trophic state) are used to model likelihood of future events within a forecasting framework. This framework uses web services to retrieve observational data for climate and hydrologic variables from external databases via web services, which are used as inputs for a predictive Bayesian Network model. This framework is designed to support monitoring and advisory decisions, with the aim of making water quality monitoring and planning strategies more proactive and efficient. The predictive model and underlying relationships between external factors and chl- *a* concentrations also add scientific value by identifying scales at which these factors have significant influence on lake water quality.

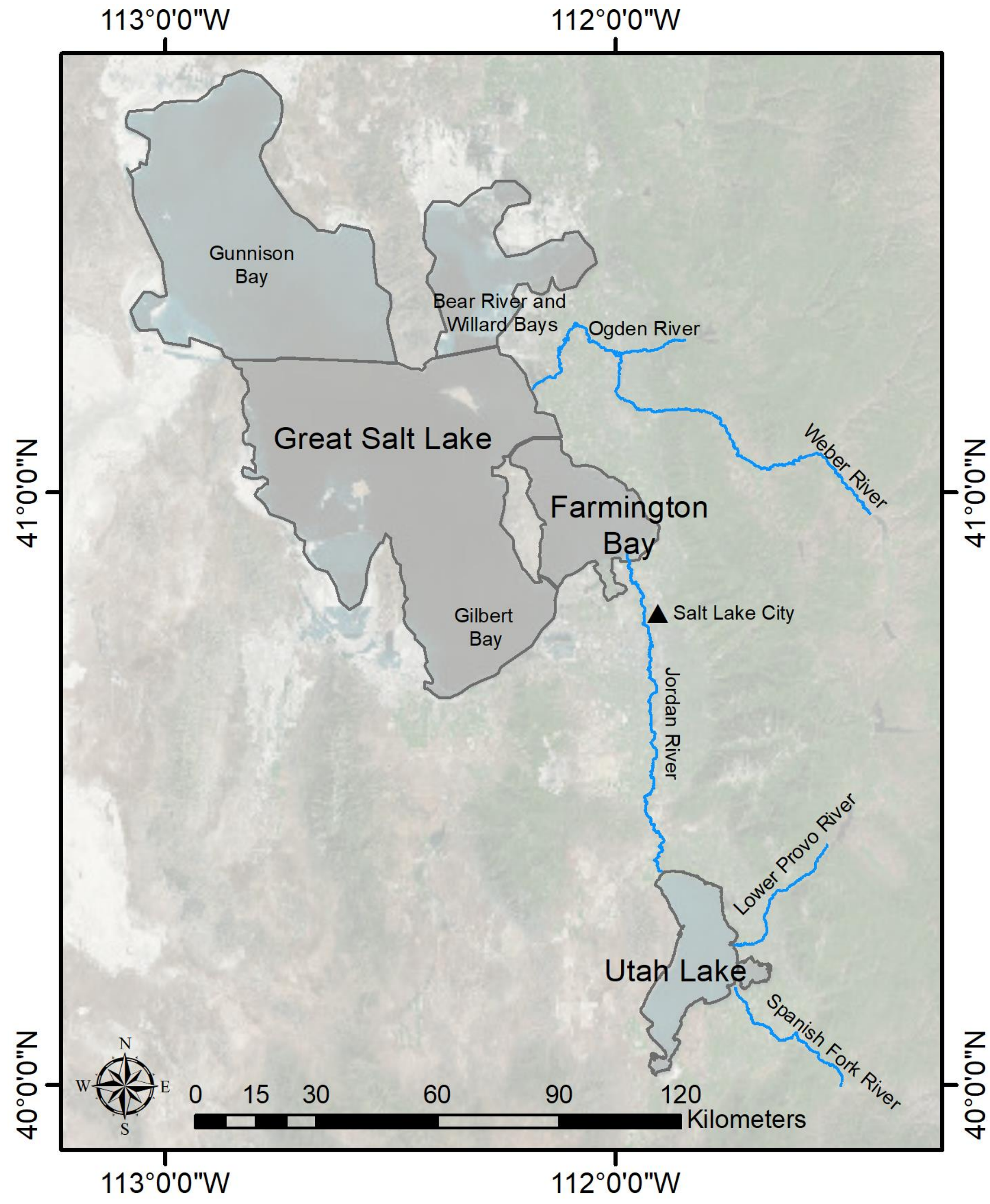

Figure 1.1 Great Salt Lake System. The focus of this research on this system is primarily on the Southern GSL, Farmington Bay, and Utah Lake.

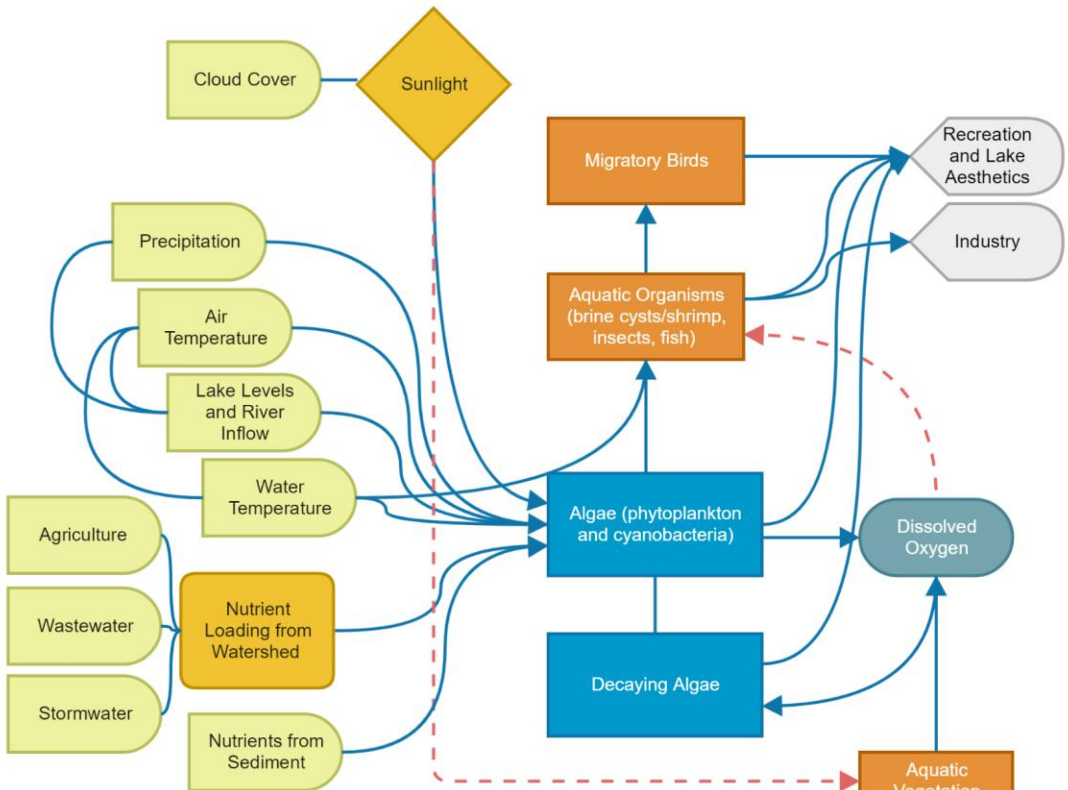

Figure 1.2 Role of Algae in the GSL Ecosystem. Blue lines indicate contribution to other components of the ecosystem, dashed red lines are used to show relationships where ecosystem components are in competition with algal blooms for resources.

Table 1.1 Review of Significant Field Sampling Studies on Algal Bloom Characteristics and Influencing Factors in the GSL System

Authors and Year Published	**Water Quality Components of Interest**	**Main Findings**	**Temporal Extent**	**Spatial Extent**
Goel and Myers 2009	Cyanobacteria/Algae speciation	Algae bloom in FB may not always be composed of cyanobacteria	Summer 2008	6 sites in Farmington Bay
Wurtsbaugh et al. 2012	Algae, nutrients, salinity, metals	Bays exhibit stark differences, including in nutrient limitation of algal growth; Salinity influences cyanobacteria growth; Factors affecting year-to-year changes are not fully understood	2002-2004,2005-2007, 2009	12 sites between Bear River Bay, Gilbert Bay, Farmington Bay
Larson and Belovsky 2013	Algae, nutrients, temperature	Species richness decreases with increasing salinity, and increases with nutrients, certain species abundance is positively affected by cooler temperatures		Laboratory treatments of water collected near Antelope Island
Wurtsbaugh 2015	Algae, nutrients, salinity, temperature	Increase in overall chl-*a* concentrations and changes in speciation with increase in salinity; Algal succession between diatoms-green algae-cyanobacteria; Salinity impacts nutrient-limiting effects	5 dates between June 2012-June 2013	9 stations in Farmington Bay
Marden 2015	Nutrients, algae speciation, abiotic factors	Spatial and temporal heterogeneity in chl-*a* and nutrient concentrations; North-South gradients; Combination of nutrients, salinity, and zooplankton grazing affect algal population size and structure	Mar-Nov 2013	8 stations within Farmington Bay, 1 station in Gilbert Bay

1.5 References

Acuña, W. C., Gonzalez, C. J., and Aqueveque, V. G. (2017). "La Chimba, Antofagasta, Chile – Oxygen depletion and hydrogen sulfide gas mitigation due to harmful algal blooms." *Harmful Algal Blooms (HABs) and Desalination: A Guide to Impacts, Monitoring, and Management*, edited by D. M. Anderson, S.F.E. Boerlage, and M. B. Dixon, Intergovernmental Oceanographic Commission, Paris, France.

Alameddine, I., Cha, Y., and Reckhow, K. H. (2011). "An evaluation of automated structure learning with Bayesian networks: an application to estuarine chlorophyll dynamics." *Environmental Modelling & Software*, 26(2), 163-172.

Andersen, M. E., Keleher, C. J., Rasmussen, J. E., Hansen, E. S., Thompson, P. D., Speas, D. W., Routledge, M. D., and Hedrick, T. N. (2007)."Status of June sucker in Utah Lake and refuges." *American Fisheries Society Symposium*, 39-58.

Anderson, D. M., Glibert, P. M., and Burkholder, J. M. (2002). "Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences." *Estuaries*, 25(4), 704-726.

Anderson, T. R. (2005). "Plankton functional type modelling: running before we can walk?" *Journal of Plankton Research*, 27(11), 1073-1081.

Anderson, D. M. (2009). "Approaches to monitoring, control and management of harmful algal blooms (HABs)." *Ocean & Coastal Management*, 52(7), 342-347.

Anderson, D. M., Cembella, A. D., and Hallegraeff, G. M. (2012). "Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management." *Annual Review of Marine Science*, 4, 143-176.

Arnow, T. (1984). "Water-level and water-quality changes in Great Salt Lake, Utah, 1847-1983." Vol. 913. US Department of the Interior, Geological Survey. US Government Printing Office.

Arnow, T., and Stephens, D. W. (1990). "Hydrologic characteristics of the Great Salt Lake, Utah, 1847-1986." United States Geologic Survey Water-Supply Paper 2332. US Government Printing Office.

Backer, L. C., McNeel, S. V., Barber, T., Kirkpatrick, B., Williams, C., Irvin, M., Zhou, Y., Johnson, T. B., Nierenberg, K., and Aubel, M. (2010). "Recreational exposure to microcystins during algal blooms in two California lakes." *Toxicon*, 55(5), 909-921.

Bioeconomics, Inc. (2012). Economic Significance of the Great Salt Lake to the State of Utah. *Report for the Great Salt Lake Advisory Council*, <http://www.fogsl.org/issuesforum/2012/wp-content/uploads/2012/05/Myers_GSLAdvisoryCouncil_Economics.pdf> (June 19, 2018)

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V. V., Morton, S. L., Perkins, D. A., Reavie, E. D., Scott, G. I., Smith, S. A., and Steevens, J. A. (2016). "Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems?" *Environmental Toxicology and Chemistry*, 35(1), 6-13.

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V. V., Morton, S. L., Perkins, D. A., Reavie, E. D., Scott, G. I., Smith, S. A., and Steevens, J. A. (2017). "In some places, in some cases, and at some times, harmful algal blooms are the greatest threat to inland water quality." *Environmental Toxicology and Chemistry*, 36(5), 1125-1127.

Carmichael, W. W., & Boyer, G. L. (2016). Health impacts from cyanobacteria harmful algae blooms: implications for the North American Great Lakes. *Harmful Algae*, 54, 194-212.

Chen, L., Delatolla, R., D'Aoust, P. M., Wang, R., Pick, F., Poulain, A., and Rennie, C. D. (2017). "Hypoxic conditions in stormwater retention ponds: potential for hydrogen sulfide emission." *Environmental Technology*, 1-12.

Cox, R.R, Kadlec, J.A. (1995). "Dynamics of potential waterfowl foods in Great Salt Lake marshes during summer." *Wetlands*, 15(1), 1-8.

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T. and Thornburgh, D.J. (2009). Eutrophication of US freshwaters: analysis of potential economic damages. *Environmental Science and Technology*, 43(1), 12-19.

Fischer, A., Ter Laak, T., Bronders, J., Desmet, N., Christoffels, E., van Wezel, A., & van der Hoek, J. P. (2017). "Decision support for water quality management of contaminants of emerging concern." *Journal of Environmental Management*, 193, 360-372.

Forio, M. A. E., Landuyt, D., Bennetsen, E., Lock, K., Nguyen, T. H. T., Ambarita, M. N. D., Musonge, P. L. S., Boets, P., Everaert, G., and Dominguez-Granda, L. (2015). "Bayesian belief network models to analyse and predict ecological water quality in rivers." *Ecological Modelling*, 312, 222-238.

Goel, R., and Myers, L. (2009). "Evaluation of Cyanotoxins in the Farmington Bay, Great Salt Lake, Utah." *Project Report*, < http://www.cdsewer.org/GSLRes/2009_CYANOBACTERIA_PROJECT_REPORT.pdf> (June 19, 2018)

Kim, H.G. (2006). "Mitigation and controls of HABs." *Ecology of Harmful Algae*, edited by E. Granéli and J.T. Turner, Springer, Berlin, Germany.

GSLEP. (2018). "OOOhhhoooh that smell!" *Research, Management and Conservation*. <http://wildlife.utah.gov/gsl/facts/smell.php> (December 15, 2015).

Hallegraeff, G. M. (1993). "A review of harmful algal blooms and their apparent global increase." *Phycologia*, 32(2), 79-99.

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., Dortch, Q., Gobler, C. J., Heil, C. A., and Humphries, E. (2008). "Eutrophication and harmful algal blooms: a scientific consensus." *Harmful Algae*, 8(1), 3-13.

Ho JC, Michalak AM. 2017. "Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading." *Journal of Great Lakes Research*, 43(3), 221-228.

Hoagland, P., & Scatasta, S. (2006). “The economic effects of harmful algal blooms.” *Ecology of Harmful Algae*, edited by E. Granéli and J.T. Turner, Springer, Berlin, Germany.

Hooton, L. W. (1989). "Utah Lake and Jordan River Water Rights and Management Plan." <http://www.slcdocs.com/utilities/PDF%20Files/utah&jordan.PDF> (June 19, 2018)

Hunter, P. (1998). "Cyanobacterial toxins and human health." *Journal of Applied Microbiology*, 84(1), 35-40.

Larson, C. A., and Belovsky, G. E. (2013). "Salinity and nutrients influence species richness and evenness of phytoplankton communities in microcosm experiments from Great Salt Lake, Utah, USA." *Journal of Plankton Research,* 35(5), 1154-1166.

Malve, O., Laine, M., Haario, H., Kirkkala, T., and Sarvala, J. (2007). "Bayesian modelling of algal mass occurrences—using adaptive MCMC methods with a lake water quality model." *Environmental Modelling & Software*, 22(7), 966-977.

Marcarelli, A. M., Mills, Michael D., Wurtsbaugh, Wayne A. (2001). "The Great Salt Lake doesn’t stink...But Farmington Bay does!" *Friends of the Great Salt Lake Newsletter*, 7(2), 5-13.

Marden, B. M., T.; Richards, D. (2015). "Factors Influencing Cyanobacteria Blooms in Farmington Bay, Great Salt Lake, Utah." *A Progress Report of Scientific Findings From the 2013 Growing Season*. Prepared For: The Jordan River/Farmington Bay Water Quality Council.

Merritt, L. B. (2017). "Utah Lake: A Few Considerations." <https://le.utah.gov/interim/2017/pdf/00004935.pdf> (July 1, 2018)

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., Chaffin, J. D., Cho, K., Confesor, R., and Daloğlu, I. (2013). "Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions." *Proceedings of the National Academy of Sciences*, 110(16), 6448-6452.

Michelakaki, M., and Kitsiou, D. (2005). "Estimation of anisotropies in chlorophyll a spatial distributions based on satellite data and variography." *Global Nest Journal*, 7(2), 204-211.

Naftz, D., Angeroth, C., Kenney, T., Waddell, B., Darnall, N., Silva, S., Perschon, C., and Whitehead, J. (2008). "Anthropogenic influences on the input and biogeochemical cycling of nutrients and mercury in Great Salt Lake, Utah, USA." *Applied Geochemistry*, 23(6), 1731-1744.

Nicholson, B., and Marcarelli, Amy (2004). "The Paradox of a Great Salt Lake." *Southwest Hydrology*, Department of Aquatic, Watershed and Earth Resources, Utah State University, 24-25.

Nojavan, F., Qian, S. S., Paerl, H. W., Reckhow, K. H., and Albright, E. A. (2014). "A study of anthropogenic and climatic disturbance of the New River Estuary using a Bayesian belief network." *Marine Pollution Bulletin*, 83(1), 107-115.

Obenour, D. R., Gronewold, A. D., Stow, C. A., and Scavia, D. (2014). "Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts." *Water Resources Research*, 50(10), 7847-7860.

Obropta, C. C., Niazi, M., & Kardos, J. S. (2008). Application of an environmental decision support system to a water quality trading program affected by surface water diversions. *Environmental Management*, 42(6), 946.

Paerl, H. W. (1988). "Nuisance phytoplankton blooms in coastal, estuarine, and inland waters." *Limnology and Oceanography*, 33(4, part 2), 823-843.

Paerl, H. W., Fulton, R. S., Moisander, P. H., and Dyble, J. (2001). "Harmful freshwater algal blooms, with an emphasis on cyanobacteria." *The Scientific World Journal*, 1, 76-113.

Paerl, H. W., and Otten, T. G. (2013). "Harmful cyanobacterial blooms: causes, consequences, and controls." *Microbial Ecology*, 65(4), 995-1010.

Page, B. P., Kumar, A., and Mishra, D. R. (2018). "A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom." *International Journal of Applied Earth Observation and Geoinformation*, 66, 69-81.

Penrod, E. (2015). "After Utah Lake-related dog deaths, experts recommend protocol for toxic algae." *Salt Lake Tribune.* Salt Lake City, UT.

Price, R. K and Vojinovic, Z. (2011). *Urban Hydroinformatics: Data, Models, and Decision Support for Integrated Urban Water Management*, IWA Publishing, London, U.K.

Rushforth, S. R., and Squires, L. E. (1985). "New records and comprehensive list of the algal taxa of Utah Lake, Utah, USA." *The Great Basin Naturalist*, 45(2), 237-254.

Smith, V. H. (2003). "Eutrophication of freshwater and coastal marine ecosystems a global problem." *Environmental Science and Pollution Research*, 10(2), 126-139.

Steffen, M. M., Belisle, B. S., Watson, S. B., Boyer, G. L., & Wilhelm, S. W. (2014). Status, causes and controls of cyanobacterial blooms in Lake Erie. *Journal of Great Lakes Research*, 40(2), 215-225.

US Congress. (2014). "Harmful Algal Bloom and Hypoxia Research and Control Amendments Act of 2014." Pub. S. 1254. Washington, DC.

Utah DEQ. (2016). “Utah Lake, Jordan River, Canals Algal Bloom 2016.” <https://deq.utah.gov/legacy/divisions/water-quality/health-advisory/harmful-algal-blooms/bloom-events/bloom-2016/utah-lake-jordan-river/index.htm>

Wang, X., and Liu, R. (2005). "Spatial analysis and eutrophication assessment for chlorophyll a in Taihu Lake." *Environmental Monitoring and Assessment*, 101(1-3), 167-174.

Whitehead, P., Wilby, R., Battarbee, R., Kernan, M., and Wade, A. J. (2009). "A review of the potential impacts of climate change on surface water quality." *Hydrological Sciences Journal*, 54(1), 101-123.

Wurtsbaugh, W. A., and Berry, T. S. (1990). "Cascading effects of decreased salinity on the plankton chemistry, and physics of the Great Salt Lake (Utah)." *Canadian Journal of Fisheries and Aquatic Sciences*, 47(1), 100-109.

Wurtsbaugh, W. A. (2008). "Nutrient loading and eutrophication in the Great Salt Lake." *Watershed Sciences Faculty Publications*, Paper 299.

Wurtsbaugh, W. A., Marcarelli, A. M., and Boyer, G. L. (2012). "Eutrophication and metal concentrations in three bays of the Great Salt Lake (USA)." *Watershed Sciences Faculty Publications*, Paper 550.

Wurtsbaugh, W. A. (2015). "Factors affecting the spatial and temporal variability of cyanobacteria, metals, and biota in the Great Salt Lake, Utah." Prepared For: Utah Division of Water Quality, Department of Environmental Quality and Utah Division of Forestry, Fire and State Lands, Department of Natural Resources.

CHAPTER 2

SPATIOTEMPORAL VARIABILITY OF LAKE WATER QUALITY IN THE CONTEXT OF REMOTE SENSING MODELS

Reprinted with permission from. Hansen, C. H., Burian, S. J., Dennison, P. E., and Williams, G. P. (2017). "Spatiotemporal Variability of Lake Water Quality in the Context of Remote Sensing Models." *Remote Sensing*, 9(5), 409.

remote sensing

MDPI

Article

Spatiotemporal Variability of Lake Water Quality in the Context of Remote Sensing Models

Carly Hyatt Hansen [1,*], Steven J. Burian [1], Philip E. Dennison [2] and Gustavious P. Williams [3]

1 Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112, USA; steve.burian@utah.edu
2 Department of Geography, University of Utah, Salt Lake City, UT 84112, USA; dennison@geog.utah.edu
3 Department of Civil and Environmental Engineering, Brigham Young University, Provo, UT 84602, USA; gus.p.williams@byu.edu
* Correspondence: carly.hansen@utah.edu

Academic Editors: Yunlin Zhang, Claudia Giardino, Linhai Li, Deepak R. Mishra and Prasad S. Thenkabail
Received: 25 February 2017; Accepted: 21 April 2017; Published: 26 April 2017

Abstract: This study demonstrates a number of methods for using field sampling and observed lake characteristics and patterns to improve techniques for development of algae remote sensing models and applications. As satellite and airborne sensors improve and their data are more readily available, applications of models to estimate water quality via remote sensing are becoming more practical for local water quality monitoring, particularly of surface algal conditions. Despite the increasing number of applications, there are significant concerns associated with remote sensing model development and application, several of which are addressed in this study. These concerns include: (1) selecting sensors which are suitable for the spatial and temporal variability in the water body; (2) determining appropriate uses of near-coincident data in empirical model calibration; and (3) recognizing potential limitations of remote sensing measurements which are biased toward surface and near-surface conditions. We address these issues in three lakes in the Great Salt Lake surface water system (namely the Great Salt Lake, Farmington Bay, and Utah Lake) through sampling at scales that are representative of commonly used sensors, repeated sampling, and sampling at both near-surface depths and throughout the water column. The variability across distances representative of the spatial resolutions of Landsat, SENTINEL-2 and MODIS sensors suggests that these sensors are appropriate for this lake system. We also use observed temporal variability in the system to evaluate sensors. These relationships proved to be complex, and observed temporal variability indicates the revisit time of Landsat may be problematic for detecting short events in some lakes, while it may be sufficient for other areas of the system with lower short-term variability. Temporal variability patterns in these lakes are also used to assess near-coincident data in empirical model development. Finally, relationships between the surface and water column conditions illustrate potential issues with near-surface remote sensing, particularly when there are events that cause mixing in the water column.

Keywords: spatiotemporal variability; water quality; chlorophyll-a; near-coincident remote sensing

1. Introduction

Over the past decade, remote sensing of water quality has become more widely used and the extent of applications has grown tremendously, especially in non-coastal environments. Notable inland water quality applications of remote sensing include large-scale quality and clarity surveys [1–4] and real-time tracking and forecasting of nuisance algal blooms (NABs) or harmful algal blooms (HABs) [5,6]. The general process of developing an empirical remote sensing model for algal blooms typically involves: downloading and processing of remote sensing imagery (which may include atmospheric correction and conversion from digital numbers to reflectance at the near-surface of the water body), collecting coincident (or near-coincident) field measurements of chlorophyll-a (or other parameters related to biomass or levels of toxins), and using regression or other statistical modeling techniques to develop a relationship between the field-measured concentrations and remotely sensed reflectance from the corresponding pixel or group of pixels. Multiple sensors offer greater coverage with varying overpass frequencies and extents, and band combinations which are more optimal for characterization of water quality conditions. Increased availability of imagery data and processed data products has also facilitated increased use and application. Despite all of these advances, there are a number of issues that remain to be addressed to support more effective and accurate remote sensing model development and application. Many of these issues stem from traditional assumptions associated with the use and application of remote sensing data, and do not consider conditions and processes that are specific to the water bodies of interest.

Water quality conditions, particularly algal growth, in lakes and reservoirs have been shown to change relatively quickly (i.e., seasonally or sub-seasonally) [7–9]. Algal bloom variability in inland waters also occurs on smaller spatial scales than in the open ocean. Spatial and temporal variability in water quality may be caused by a number of processes, such as resuspension of suspended sediments and point-source inflow of nutrients [10]. Increased variability in lake and reservoir water quality requires that in situ data used to develop remote sensing water quality models represent conditions at the time of the imagery acquisition–to the extent possible. Often, the historical records do not provide exact temporal matches between the in situ samples and the satellite overpass, requiring the use of "near-coincident" data, or some relaxation of a definition of a "match." Coastal and lake water clarity and quality remote sensing literature report a wide range of time-windows for considering data to be near-coincident. Reported windows range from $\pm$3 h [11], same day [12], one day [4,13], seven days [2,14], to $\pm$10 days [1] between the satellite image acquisition and the field samples used for calibration. Often, a particular time-window for near-coincident matches is arbitrarily chosen (e.g., using an arbitrary increase in the percentage of samples that match with a satellite image [15]), or the study states that the relaxation of the time-window improved the model fit, without detailing the actual improvement [1].

Another issue that is often overlooked in water quality remote sensing applications is thorough review and evaluation of appropriate sensors in the context of a specific water body (which has unique spatial and temporal characteristics). Sensor characteristics can have large implications for the utility of the resulting model and dataset. Model application determines the sensor choice and could depend on a number of factors: the spatial resolution (which is limited by the size of the water body or multiple waterbodies in a region), the spatial variability within the water body, the desired return time (which is influenced by the temporal variability of the water quality processes), the length of historic record, spectral resolution (which determines the ability of the sensor to discriminate or more accurately determine conditions and which parameters can be estimated), the available processing resources (from the imagery data and data products to the personnel who will perform data processing and analysis), and the scope of the application (both spatial and temporal). For empirical model development, information from the field (e.g., concentration of chlorophyll-a measured at a single point on the water body) is matched to information from the satellite (reflectance averaged over a single pixel or group of pixels). Therefore, the spatial variability of the water body may influence the choice of satellite. For example, if the algae concentrations vary substantially on the order of 20–40 m, then a satellite with a resolution of 30 m will be sufficient, while a satellite with a resolution of 500 or 1000 m would be too coarse to adequately represent the variability of the chlorophyll concentrations. One review suggests different medium spatial resolution satellites (e.g., Landsat) and coarser spatial resolution satellites (e.g., MODIS) for water clarity and quality studies be selected based primarily on the size of the water body [16], however, other characteristics of the lake, namely the ability of different spatial resolutions (e.g., Landsat resolution of 30 m or SENTINEL-2 resolution of 10–60 m compared to MODIS resolution of 250–1000 m) to represent spatial variability within the lake or the ability of more frequent overpasses to address temporal variability (e.g., Landsat every 16 days compared to SENTINEL-2 every 5 days and MODIS every 1–2 days) are not considered.

Finally, remotely sensed data are limited by the optical depth of the water column (the depth at which light is able to penetrate), which means that the estimates are limited to near-surface algae populations. Optical depth is also a function of chlorophyll concentration; as the near-surface algae populations increase, optical depth decreases. However, algae thrive not only at the surface but exist throughout the water column. Algal population characteristics (species, diversity, etc.) may vary with depth, especially when the water column is stratified and there are differences in oxygen or salinity [17,18]. Concerns have been raised about the utility of only sensing and estimating the surface of the lake given these variable conditions throughout the water column. It is therefore important to explore the relationship between surface and water-column algae concentrations and the variability within the water column when evaluating the limitations of remotely-sensed surface estimates.

This study uses field measurements of chlorophyll to evaluate techniques and assumptions that are often used in remote sensing models of algae and surface water quality. While there are many additional considerations for water quality (particularly algae/chlorophyll concentrations) this paper focuses on the three issues outlined above: (1) selecting sensors which are suitable for the spatial and temporal variability in the water body; (2) determining appropriate uses of near-coincident data in empirical model calibration; and (3) recognizing potential limitations of remote sensing measurements which are biased toward surface and near-surface conditions.

Study Area

The study area for this paper is the Utah Lake and Great Salt Lake (GSL) system. This lake system is important for recreation and ecosystem services for the urban areas that are concentrated in the hillsides and valleys to the east of these lakes. During the summer of 2016, Utah Lake and Farmington Bay of the GSL experienced massive cyanobacterial algal blooms. While large algal blooms in these lakes are not particularly rare, the rapid development and magnitude of the recent blooms spurred widespread attention and motivated increased interest in monitoring these waters, particularly through remote sensing because the size of the lakes make them difficult to monitor through field sampling alone. Data were collected with water quality sondes at a number of locations throughout the system (shown on the map in Figure 1) throughout the summer of 2016 to support this research.

Previous studies in the Utah Lake and GSL system have explored variation in algal speciation throughout the growing season and environmental factors which contribute to species diversity [19–22]. Historical sampling campaigns on Utah Lake revealed typical algal succession, with diatoms and then green algae dominating in early summer, and then cyanobacteria dominating during the late summer months, and a general decrease in species diversity throughout the summer [21,22]. In Farmington Bay and the GSL, studies have focused on speciation and presence of toxins in cyanobacteria. These studies have found seasonal trends in algae growth and have observed stark differences between algae types in different regions of the GSL and Farmington Bay [19,20,23,24]. These studies improve understanding of the algae populations in this lake system; however, they lack important information about spatial or temporal variability at scales that are necessary for improving remote sensing model development.

The Great Salt Lake is divided roughly in half by a railroad causeway which runs East-West, separating the much more saline (roughly 28% salinity) North Arm, which includes Gunnison Bay and Bear River/Willard Bay, from the South Arm (Gilbert Bay and Bridger Bay) and Farmington Bay, which is further separated by an automobile causeway. These bays maintain a salinity between 11% and 15% [25] and at the north end of Farmington Bay, salinity is typically around 8% [20]. These lakes are relatively shallow, with an average depth of approximately 4.2 m in Gilbert Bay and an average depth of approximately 1 m in Farmington Bay. Secchi depth (as a measure of transparency) ranges between 2 and 5 m in the South Arm of the GSL, while in Farmington Bay, it is regularly less than 0.3 m [26]. Utah Lake, which flows into the Great Salt Lake through the Jordan River is also a shallow lake (average depth of 2.74 m) and while it is a freshwater lake, it has high dissolved solids, resulting in slightly saline conditions [27]. High rates of suspended sediments result in high turbidity, and prior to the large algal bloom in 2016, the Secchi depth in the middle of Utah Lake was roughly 0.2 m.

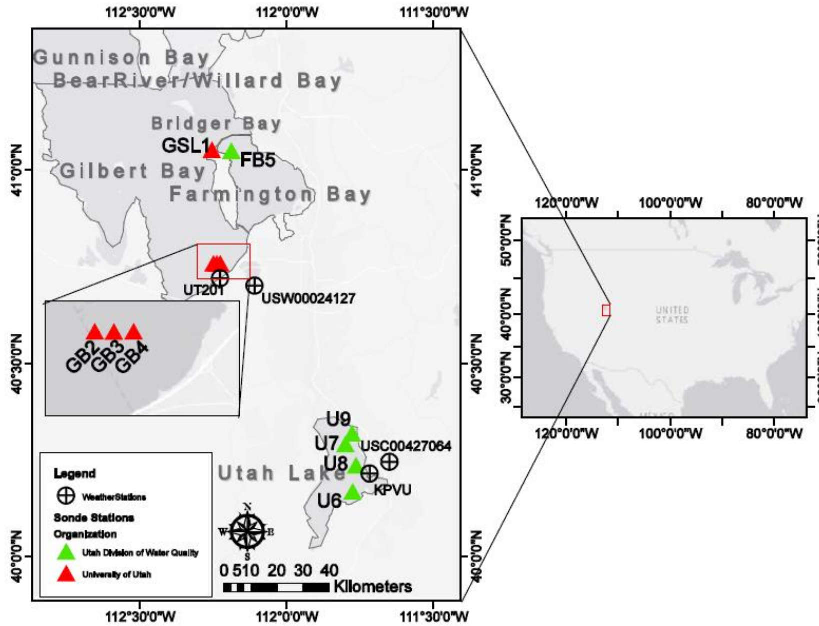

Figure 1. Sampling Locations and Study Area.

2. Materials and Methods

2.1. Data Collection

The collection of water quality samples was designed to provide information about algae biomass (measured as chlorophyll-a) and its: (1) temporal variability (through repeated sampling visits and high-frequency sampling); (2) spatial variability (through multiple sites and/or offsets); and (3) surface–water column relationships. Chlorophyll-a data were collected by researchers at the University of Utah (U of Utah) using a Hydrolab DS5 (OTT Hydromet) multiparameter sonde equipped with a submersible fluorescence Chlorophyll-a sensor (range of 0.03–500 µg/L). Chlorophyll-a data were also provided by the Utah Division of Water Quality (UDWQ) measured using YSI EXO 2 multiparameter sonde (with submersible fluorescence Chlorophyll-a sensor (range of 0–400 µg/L) coupled with a Nexsens CB-450 buoy platform. Sampling locations were chosen based on accessibility. During the study period, low water levels, exposed reef-like bioherms, and deep sediments restricted boat and individual access to many locations in the lakes that may otherwise have been sampled. Details of the sampling at each station are summarized below and in Table 1, including the duration of sampling periods and the types of samples collected. Durations and frequencies of data collection were determined by the availability of equipment and personnel, and local weather conditions. Data collected by the University of Utah are shared under the Creative Commons Attribution CC BYU License [28] and data collected by the UDWQ are available through the iUTAH Time Series Analyst data portal.

Table 1. Summary of Data Collection Periods and Methods.

Lake	Stations	Organization	Sampling Periods (2016)	7.5 m Offsets	Surface (<1 m)	Water Profiles	Approximate Lake Depth During Study Period (m)
Main GSL	GSL1	U of Utah	23–31 July	X	X	-	0.8
	GB2; GB3; GB4	U of Utah	6–16 June; 6–14 July; 12–22 Aug	X	X	X	5.1
Farmington Bay	FB5	UDWQ	8 July–28 July	-	X	-	0.5
Utah Lake	U6; U7; U8	UDWQ	28 Aug–13 Sept	-	X	-	1–1.5
	U9	UDWQ	15 July–8 Aug	-	X	-	1–1.5

2.1.1. UDWQ Data

UDWQ sondes were installed in a variety of locations in Utah Lake and Farmington Bay following the large July 2016 algal blooms. The site names for these sites have been modified to maintain consistency with the naming convention of the University of Utah sites. One temporary fixed sonde was placed approximately 0.75 m below the surface at station U9 (UDWQ Site 4917310) in Utah Lake, providing daily measurements between 15 July and 8 August, 2016. The sondes in stations U6 (UDWQ Site 4917390), U7 (UDWQ Site WVineyard), and U8 (UDWQ Site WProvo) were installed on buoys anchored at the locations shown in Figure 1, and provided daily measurements at approximately 0.3 m below the surface between 28 August and 13 September 2016. Water depths in Utah Lake during this time period were between 1 and 1.5 m. Finally, a fixed sonde in Farmington Bay at station FB5 (UDWQ Site 4895200) provided daily measurements between 8 July and 28 July, 2016 at a depth of approximately 0.3 m below the surface (due to extremely low water levels, which were approximately 0.5 m at this time). The measurements for these sondes (which were reported at a 15-min frequency) were averaged between 11:00–11:30 a.m. in order to maintain consistency in day-to-day comparisons (reducing the effect of diurnal patterns of algae on the chlorophyll measurements which peaks during midday and then drops in the evening). These daily measurements were used in exploring temporal variability.

2.1.2. University of Utah Data

While the fixed UDWQ sondes in Utah Lake and Farmington Bay provide stationary data for exploration of temporal variability, data collection by the University of Utah was designed to explore temporal variability as well as variability on different spatial scales. Data collected by the University of Utah was focused in the main body of the South Arm of the GSL (Gilbert Bay and Bridger Bay). Surface data at the Gilbert Bay sites were consistently collected between 9:00 and 11:30 a.m. (again, to minimize the effects of diurnal patterns of photosynthesis). Data collection took place during three periods: 6, 8, 9, 10 and 13 June; 6, 7, 8, 12 and 14 July; and 12, 15, 16, 17 and 22 August. At these sites (GB2, GB3 and GB4), approximately 20–30 measurements were taken at a 1-s frequency at an average depth of 0.4 m below the surface and averaged. The Gilbert Bay sites (prefixed with GB) which were navigable by boat, were located approximately 1000 m apart, which is the same scale as the coarsest MODIS spatial resolution. At each of these sites, data were also collected at offsets to the site center to represent sub-Landsat and sub-SENTINEL-2 resolution. These offset samples were spaced at approximately 7.5 m increments (i.e., 7.5, 15, 22.5 and 30 m) from the original sites GB2, GB3 and GB4. The offsets were identified with suffixes a, b, c and d, so that the first offset (7.5 m) from GB2 was identified as GB2a, the second offset (15 m) from GB2 was GB2b, etc.) At these sites, lake current and wind patterns differed from one sampling day to the next, resulting in variable drift directions between the GB sites and their offsets, though it was generally consistently in the southwest direction. Nonetheless, relative distances between the original sites and the offsets were maintained. Approximately 20–30 measurements at the GSL1 site were collected at a 1-s frequency approximately 0.3 m below the surface and averaged in a July sampling period (23, 24, 27, 30 and 31 July). Data collection at this site also included sampling at offsets at the same increments (7.5, 15, 22.5 and 30 m) east of the original site.

The data at Bridger Bay were averaged at approximately 0.3 m below the surface (due to low lake levels at this location), and were consistently collected in the afternoon (due to equipment availability and to reduce effect of diurnal patterns).

In addition to the surface data obtained at the Gilbert Bay sites, measurements were collected throughout the water column to examine relationships between chlorophyll measurements at different depths. At sites GB2, GB3, and GB4, data were collected over the water profile, by manually lowering the sonde at approximately 0.3 m/s and recording at a 1-s frequency. Profiles were created by averaging the concentrations over 1 m intervals from 0–6 m) to represent different ranges of the water column.

For the sites reached by boat, we approached the locations from the opposite direction of the lake current and turned off the engine, allowing the boat to drift to the sites and offsets in an effort to reduce the amount of artificial mixing caused by the engine. Despite these efforts, some amount of mixing from the engine may have occurred which would have an effect on the measured concentrations and subsequent variability, particularly near the surface. The FB site and offsets were reached by foot, and mixing may have been caused by stirring up sediments.

2.1.3. Meteorological Data

In order to examine conditions that may contribute to surface mixing in the lakes, meteorological data were collected from MesoWest weather stations located near the Gilbert Bay sampling locations (Site UT201, at 40.72255, −112.22569) and near Provo Bay in Utah Lake (Site KPVU, at 40.21667, −111.71667). Parameters including wind speed (kilometers per hour) and peak wind gust (kilometers per hour) were recorded at 10 min intervals for UT201 and at 5 min intervals for KPVU. Wind speed is averaged over a daily scale and the daily peak wind gust is the maximum peak wind gust. Daily precipitation data totals (mm) and maximum temperatures (degrees Celsius) were obtained from NOAA Stations USW00024127 at 40.7034, −112.109 and USC00427064 at 40.2458, −111.6508. Comparable meteorological data near the Farmington Bay site were not available for study period.

2.2. Statistical and Graphical Analysis

To evaluate the variation over time, we computed the autocorrelation function or estimates of autocovariance [29]. These estimates were calculated for each site with regular daily sampling (all of the UDWQ sites in Utah Lake and Farmington Bay) using the "acf" function, which is built in to the R statistical software [30]. At each of the lags for these sites, we tested for statistically significant autocovariance of surface chlorophyll measurements. The autocorrelation function could not be computed for the main GSL sites (GB and GSL), since these data were not collected at regular intervals, and there were insufficient points for alternative analyses (e.g., constructing a temporal variogram). Instead, for these sites, temporal variation was analyzed graphically by calculating the difference in chlorophyll measurements between subsequent samples (for short-term variation), as well as the mean and standard deviation for each of the sampling periods (for seasonal variation).

We also examined spatial variation of surface chlorophyll concentrations with respect to the spatial resolutions of several commonly-used sensors. As noted, the distances between sites and offsets for the samples are representative of the spatial resolution of Landsat/SENTINEL-2 and MODIS band regions. The observed differences in measurements between the offsets and the sites offer insight into fine-scale variability (<30 m) that would occur at the sub-Landsat and SENTINEL-2 spatial resolutions and coarser-scale variability (1000 m) that corresponds with the spatial resolution of MODIS. To evaluate the differences between offsets, we calculated the difference and percent differences in surface measurements between the sites and their respective offsets using Equations (1) and (2):

$$Difference = Chl_{x,j} - Chl_{y,j} \tag{1}$$

$$Percent\ Difference = \left(\frac{Chl_{x,j} - Chl_{y,j}}{Chl_{x,j}}\right) * 100 \tag{2}$$

where *Chl* is the mean chlorophyll concentration between 0 and 1 m below the surface for the sampling date *j* at site *x* (e.g., GB2) and corresponding offset *y* (e.g., GB2a, GB2b, etc.).

Finally, we used linear regression to evaluate relationships between conditions at the surface and throughout the water column for the GB2, GB3, and GB4 sites for each of the sampling periods. Due to extremely low lake levels in Farmington Bay, Utah Lake, and Bridger Bay, samples at multiple depths were not possible at these locations. The regressions follow the general form of Equation (3):

$$Chl_{x,k} = m \cdot Chl_{x,l} + b \tag{3}$$

where *Chl* is the mean chlorophyll concentration at site x, at depth k below the surface, and l is the depth of 0–1 m below the lake surface. The strength of the relationship is measured through the correlation coefficient, or R^2. For this case, the correlation coefficient translates to the amount of variance at intermediate depths that is explained by the surface measurements.

3. Results

3.1. Temporal Variability

The results of the autocorrelation function are visualized in a correlogram, showing the autocorrelation of surface chlorophyll values versus the lag (days). The correlograms for each of the sites with daily sampling, shown in Figure 2, graphically illustrate how the time series is correlated with itself, or how similar measurements are from one day to measurements from some lagged time period.

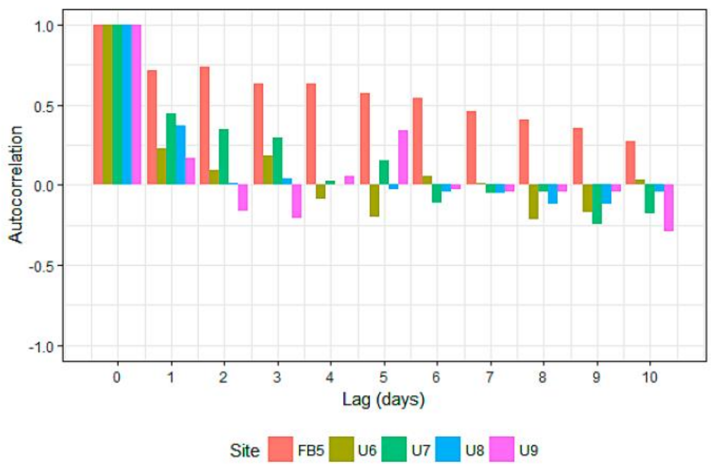

Figure 2. Autocorrelation for Utah Lake (U6, U7, U8 and U9) and Farmington Bay (FB5) Sites.

The null hypothesis, which is tested at each lag, is that there is no autocorrelation between the lagged samples. The different patterns of autocorrelation in Figure 2 show that there are major differences in the temporal autocorrelation in different parts of the lake system. At $\alpha = 0.05$, there is no statistically significant autocorrelation for all time lags for Utah Lake sites U9 and U6, and near-statistically significant autocorrelation for a lag of one day for U8 and U7. The rapid decrease in autocorrelation for many of the Utah Lake sites is evidence of high short-term variability in this body. In clear contrast with the patterns observed in Utah Lake, there is significant autocorrelation for all lags up to 11 days for the site in Farmington Bay (FB5).

For sites where it was not possible to calculate an autocorrelation function, the differences in chlorophyll measurements between subsequent samples for each of the sampling periods are shown in Figure 3.

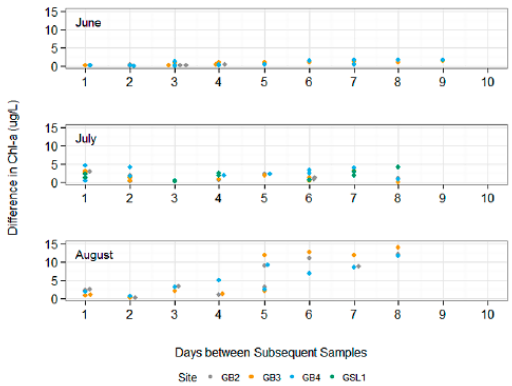

Figure 3. Temporal Variation between Subsequent Samples by Sampling Periods at the GB and GSL Sites.

In the samples from June and July, there is relatively small variation (<2 and 5 μg/L, respectively), even at 8 and 10 days between subsequent samples. In August, however, the data show a clear positive trend of increasing differences between surface chlorophyll measurements, that is, the difference between the subsequent samples increases as time between the samples increases. The data also show the variation in between subsequent measurements increases throughout the summer season. For example, in June, the mean difference at seven days between subsequent samples is 1.02 μg/L, while the mean differences in July and August at seven days are 3.05 μg/L and 9.67 μg/L, respectively. This seasonal increase in variability is also evident in comparisons of the standard deviation of surface measurements during each sampling period, shown in Figure 4. There was also a general positive trend in chlorophyll concentrations throughout the sampling period (meaning that both magnitudes of chlorophyll and variance increased throughout the summer).

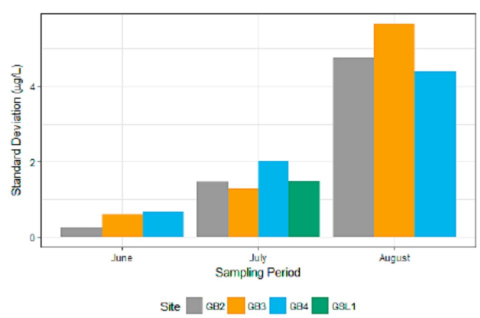

Figure 4. Standard Deviation for Surface Chlorophyll at GB and GSL Sites by Sampling Period.

3.2. Spatial Variability

To illustrate the differences in spatial resolution of several commonly-used sensors, Figure 5 compares the coverage of a portion of the study area (Utah Lake) with resolutions ranging from 30 m (Landsat 8, Band 2, 19 July 2016), to 60 m (SENTINEL-2, Band 1, 22 July 2016) and 500 m (MODIS, Band 3, 20 July 2016).

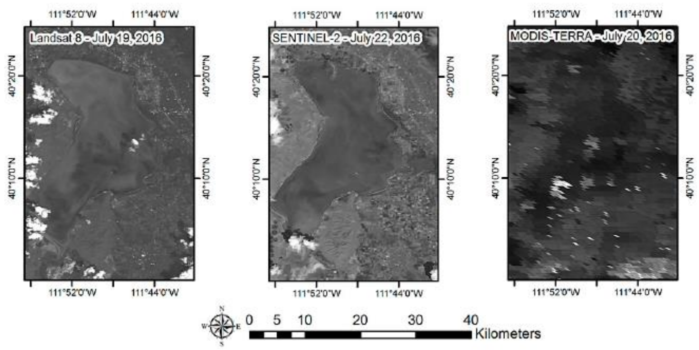

Figure 5. Comparison of Spatial Resolution in Coverage of Utah Lake at 30 m (Landsat 8, Band 2), 60 m (SENTINEL-2, Band 1) and 500 m (MODIS, Band 3).

The resolutions of Landsat and SENTINEL-2 show clear definition between the lake and the shore, and variability in surface conditions (including the extent of the large algal bloom) can be detected at both these scales. On the other hand, the coarse resolution of the MODIS image makes it difficult to delineate the shoreline and while there is some variability between the in-lake pixels, the extent of the bloom is difficult to distinguish. In the GB sites, surface chlorophyll data collected at sites and offsets correspond with the range of spatial scales for these sensors. The differences in surface chlorophyll for fine spatial scales (corresponding with Landsat/SENTINEL-2) and coarse spatial scales (corresponding with the coarsest resolution of MODIS, 1000 m) are shown in Figures 6 and 7.

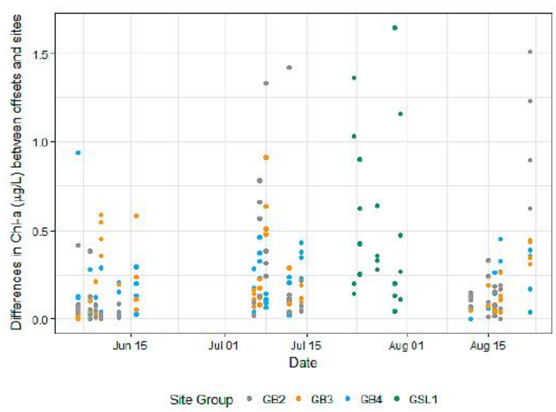

Figure 6. Variability between Sites and Offsets (<30 m distances or Sub-Landsat/Sub-SENTINEL-2 Scales) in the Great Salt Lake (GB and GSL Sites).

For site groups (where each site group includes the site and its offsets) GSL1, GB2, GB3 and GB4, there was generally less than 30 percent difference between the surface measurements at the offsets and those at the site. The plots show that the highest differences between the offsets and the sites occur in the later summer months, while relatively small differences are observed in early summer. Throughout the entire season, the maximum difference in magnitude between a site and its offsets is 1.7 µg/L.

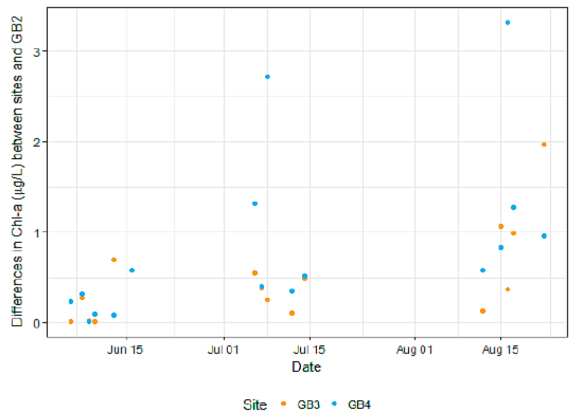

Figure 7. Variability between Sites (Approximately 1000 m distances, or MODIS Scale) in the Great Salt Lake (GB Sites).

This figure shows that even at this larger scale, the differences are still generally small (below 30 percent), though the actual difference in magnitude was higher (with a maximum difference of 3.4 µg/L) than those at the sub-pixel distances on the Landsat/SENTINEL-2 scale. Again, greater differences are observed in later summer months compared to early summer.

3.3. Surface/Water Column Measurements

The linear relationships between average surface measurements (0–1 m below the surface) and various depths (1–2 m, 2–3 m and 3–4 m) from data collected in Gilbert Bay (where water depths allowed for water column measurements) are shown in Figure 8.

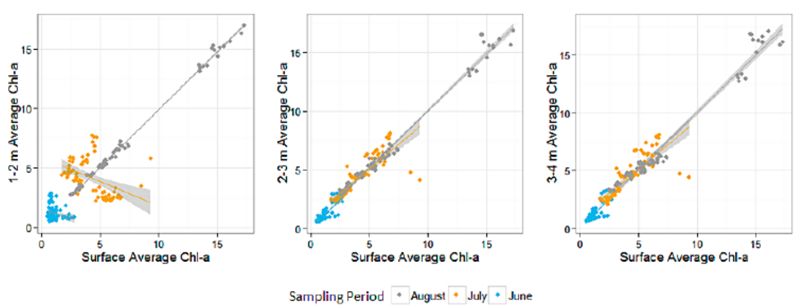

Figure 8. Relationships between Surface and Depths throughout the Water Column for GB Sites.

For 1–2 m, the overall (across all sampling periods) R^2 is 0.79; for 2–3 m it is 0.97; and for 3–4 m it is 0.96. However, the relationship is highly dependent on the sampling period, particularly at depths of 1–2 m. For June and July, there are virtually no relationships between the surface chlorophyll and chlorophyll at 1–2 m below the surface, and the relationships at other depths are weaker for these sampling periods than for the August sampling period.

3.4. Meteorological Record

Short-term weather events such as rainfall and high wind events have the potential to cause surface mixing and subsequently affect the observed temporal and spatial variability patterns, as well as conditions throughout the water column. Records of the daily average values for wind speed, peak daily wind gust, total daily precipitation and maximum temperature are shown for two weather stations near the Great Salt Lake and Utah Lake are shown in Figure 9.

During the periods of data collection for Utah Lake sites, conditions were relatively stable with respect to precipitation and temperature. The extremely shallow lake was likely heavily influenced by the wind, allowing for a great deal of mechanical mixing to occur. This corresponds with the low autocorrelation values in the Utah Lake sites. Other seasonal patterns in variability, such as the general increase in concentrations observed in the GB sites, correspond with the fairly stable and favorable weather conditions (lack of any large precipitation events during the mid-summer months, sustained high temperatures in late July, and a steady cooling through August).

The seasonality of the surface/water column relationship may be partially explained by weather conditions and short-term events, such as the variable temperature in June and July, and the slightly higher wind and precipitation events in the GSL in June. It is important to note that poor correlations between surface and 1–2 m depths may also be influenced by mechanical mixing caused by turbulence from the boat, which could create artificially high variability near the surface.

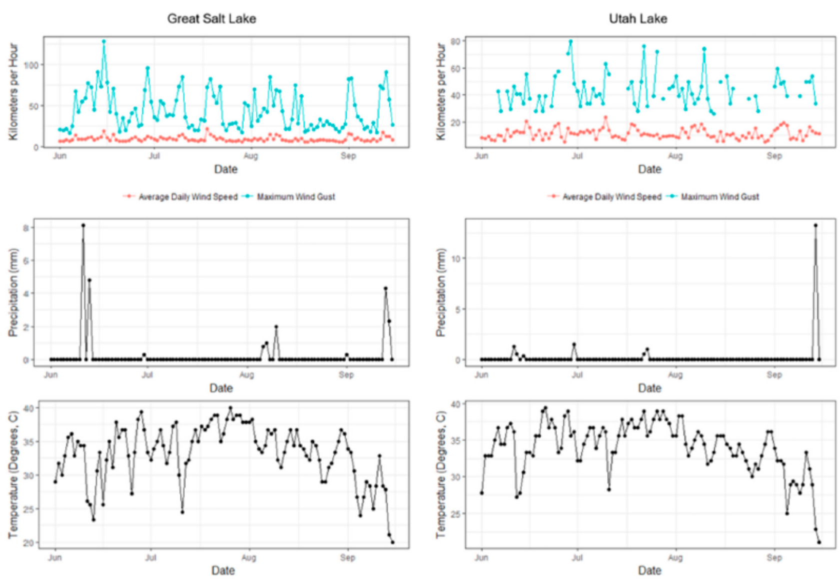

Figure 9. Daily Wind, Precipitation, and Temperature Records near the Great Salt Lake (GSL) and Utah Lake over the Period of Data Collection.

4. Discussion

The measures of variability over time (including autocorrelation, magnitude of differences between subsequent samples, and standard deviation for different sampling periods) suggest that the water bodies in the Great Salt Lake system have distinct temporal characteristics. These characteristics have important implications for remote sensing modeling techniques. The Utah Lake samples showed non-significant autocorrelation after one day, while the Farmington Bay samples showed statistically significant autocorrelation for up to 11 days. This indicates that the Utah Lake conditions are much more variable than those in Farmington Bay, with Utah Lake variation on a daily scale, rather than the near-weekly scale exhibited in Farmington Bay. In a remote sensing context, this means that shorter time windows may be needed for calibrating Utah Lake models, while longer time windows may be justified for Farmington Bay models. In the GB and GSL locations, where sampling frequencies were irregular, there was a general trend of increasing differences in chlorophyll concentrations as the time between samples increased. These differences and the overall variation increased throughout the summer, indicating that the temporal correlation may not be stationary, but decreases throughout the growing season. This increase in variability could justify a shorter time-window for near-coincident data in the later summer months than the earlier summer months.

The observed temporal patterns provide additional information for evaluating suitability of the Landsat, SENTINEL-2, and MODIS sensors for this lake system. For example, events in Utah Lake may be completely missed by the revisit time of Landsat sensors, requiring the use of multiple sensors to adequately capture the rapidly changing conditions and acknowledgment of the limitations of the temporal resolution of this sensor and its ability to describe short-term changes.

The comparisons of surface measurements between the GB and GSL sites and offsets as well as among sites were also useful in evaluating different spatial resolutions of commonly-used sensors. The relatively small variation between sites and offsets indicates that there is low variability over the distances measured by a single pixel for Landsat/SENTINEL-2 or MODIS. This suggests that these platforms, or others with similar spatial resolution, are suitable for monitoring the main body of the GSL. These results also suggest that finer spatial resolution products (such as those obtained by airborne sensors) would not necessarily provide significantly more information for this part of the system.

Finally, the linear models between concentrations at the surface and those at different depths in the water column in the GB sites show that these relationships are both depth and seasonally dependent. This result is interesting because it shows a stronger relationship between the measurements at the surface and greater depths (2–3 and 3–4 m) than between the surface and subsurface (1–2 m) measurements. If the data are analyzed by sampling period, the relationship between the surface data and the 1–2 m data exhibit a relatively strong fit for August, but not in June or July. The data at greater depths, however, exhibit relatively strong relationships during all of the sampling periods. The high variability observed at the surface and near-surface depths indicates that surface-biased estimates may be influenced by short-term weather events or human activity that causes mixture. The strong linear relationships for the other depths and for 1–2 m depths during August suggest that near-surface estimates provided by remote sensing may be strongly correlated with conditions throughout the water column, especially during periods of low surface mixing. In summary, the different relationships between surface and water column conditions highlight that surface conditions do not always reflect the conditions throughout the water column, and that the mechanical mixing processes which are unique to each water body should be taken into account before assuming any relationship between surface and water column conditions.

The spatial and temporal patterns observed in these lakes add to previous observational studies in these lakes which have focused largely on speciation and the diversity of algal populations. As species diversity decreases throughout the summer, the observations in this study also show that overall algae biomass magnitudes and variability in algae biomass increases. This relationship has both positive and negative implications for remote sensing; it provides additional motivation for using remote sensing methods during the late summer months when conditions are highly variable and more likely to be worse than early summer months, but it also highlights potential challenges associated with remote sensing of conditions when there is high species variability (leading to greater potential variability in the spectral signature of the surface waters).

5. Conclusions

The observations and analysis provided valuable insights into the Utah and GSL lake systems; however, it is important to acknowledge that the results may not be representative for all portions of the system. In particular, the surface/water column analyses in the lower portion of the GSL are not representative of the surface water/water column relationship in Utah Lake. Utah Lake is consistently much more turbid than the southern arm of the GSL, in general is shallower, and has far different mixing patterns. We recommend that this kind of analysis should be conducted in areas where unique or localized hydrodynamic disturbances exist (such as elevated exposure to wind and surface mixing, or near outfalls from wastewater treatment plants or streams where there may be increased mixing or stirring up of bottom sediments).

The temporal and spatial analysis presented in this study supports development of specific methods for future remote sensing work in this region. This support includes selecting appropriate sensors and defining appropriate time-windows for using near-coincident data. The seasonal differences in temporal correlation (as inferred by differences between subsequent samples) suggest the use of a shorter time-window for near-coincident data in calibrating empirical models in the late summer season than in the earlier summer months. We recommend that for modeling development in the main body of the GSL, near-coincident matches be limited to ± 2 days, though more relaxed time-windows could be used for early summer matches. Based on the autocorrelation of the samples in Utah Lake and Farmington Bay, we recommend limiting the time windows for considering near-coincident matches to ± 1 day for Utah Lake, while Farmington Bay may use a more relaxed time window.

Our spatial analysis showed small variations between offsets and sampling sites, indicating that Landsat/SENTINEL-2 resolution and MODIS resolutions would be appropriate for the southern arm of GSL, while finer-scale resolutions may be unnecessary as there is little variation at these smaller scales. As with the surface/water column analysis, this type of sampling in other parts of the lake system would be helpful in determining the most appropriate methods based on their unique spatial variability characteristics. From a temporal standpoint, the Landsat return time of 16 days is offset by the fact that there are multiple sensors which may be used, for example both Landsat 5 and 7 provide data for historical applications, while Landsat 8 and SENTINEL-2 provide data for more recent and ongoing applications (from 2013 and 2015, respectively). These instruments provide imagery on a more frequent basis (assuming no interference from cloud cover). However, our temporal analysis of the sensor data in Utah Lake and the main body of the GSL, shows that lake conditions change on shorter periods, and this revisit frequency may miss important changes in surface algae conditions. This is contrasted by Farmington Bay, where the conditions do not change as drastically over these time scales.

The information about spatiotemporal patterns should be considered along with other factors including: the spectral resolution of the sensors and how well the spectral measurements can describe the measures of algal biomass in certain lake environments [31], data availability (both field samples and imagery), and the historical scope (which may restrict the types of sensors which can be used) in order to meet the needs of the specific region of interest and the application. While focused on the GSL region and its unique characteristics, this study demonstrates a number of sampling and analysis techniques that could be applied in other settings to inform and improve the design of remote sensing studies. Information about the unique spatial and temporal variability patterns in a water body should be incorporated into the process of remote sensing model development, to help guide modeling decisions and assumptions.

Acknowledgments: This article was developed under Assistance Agreement No. 83586-01 awarded by the U.S. Environmental Protection Agency to Michael Barber. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. Additional funding was provided by the Department of Civil and Environmental Engineering at the University of Utah. The authors would like to acknowledge Marshall Baillie and others at the Utah Division of Water Quality for sharing data.

Author Contributions: S.B., P.D. and G.W. provided input on study concept, advising on data analysis and edits to manuscript drafts. C.H. performed data collection and analysis, and prepared the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Olmanson, L.G.; Bauer, M.E.; Brezonik, P.L. A 20-year Landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sens. Environ.* **2008**, *112*, 4086–4097. [CrossRef]
2. Kloiber, S.M.; Brezonik, P.L.; Olmanson, L.G.; Bauer, M.E. A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sens. Environ.* **2002**, *82*, 38–47. [CrossRef]
3. Sayers, M.; Fahnenstiel, G.L.; Shuchman, R.A.; Whitley, M. Cyanobacteria blooms in three eutrophic basins of the great lakes: A comparative analysis using satellite remote sensing. *Int. J. Remote Sens.* **2016**, *37*, 4148–4171. [CrossRef]
4. 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 Reserv. Manag.* **2015**, *31*, 225–240. [CrossRef]
5. Wynne, T.T.; Stumpf, R.P.; Tomlinson, M.C.; Fahnenstiel, G.L.; Dyble, J.; Schwab, D.J.; Joshi, S.J. Evolution of a cyanobacterial bloom forecast system in western Lake Erie: Development and initial evaluation. *J. Gt. Lakes Res.* **2013**, *39*, 90–99. [CrossRef]
6. Glasgow, H.B.; Burkholder, J.M.; Reed, R.E.; Lewitus, A.J.; Kleinman, J.E. Real-time remote monitoring of water quality: A review of current applications, and advancements in sensor, telemetry, and computing technologies. *J. Exp. Mar. Biol. Ecol.* **2004**, *300*, 409–448. [CrossRef]
7. Goldman, C.R.; Jassby, A.D.; Hackley, S.H. Decadal, interannual, and seasonal variability in enrichment bioassays at Lake Tahoe, California-Nevada, USA. *Can. J. Fish. Aquat. Sci.* **1993**, *50*, 1489–1496. [CrossRef]
8. Jiang, Y.-J.; He, W.; Liu, W.-X.; Qin, N.; Ouyang, H.-L.; Wang, Q.-M.; Kong, X.-Z.; He, Q.-S.; Yang, C.; Yang, B. The seasonal and spatial variations of phytoplankton community and their correlation with environmental factors in a large eutrophic Chinese lake (Lake Chaohu). *Ecol. Indic.* **2014**, *40*, 58–67. [CrossRef]
9. Wynne, T.T.; Stumpf, R.P. Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in western Lake Erie from 2002–2014. *Toxins* **2015**, *7*, 1649–1663. [CrossRef] [PubMed]
10. Mouw, C.B.; Greb, S.; Aurin, D.; DiGiacomo, P.M.; Lee, Z.; Twardowski, M.; Binding, C.; Hu, C.; Ma, R.; Moore, T. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sens. Environ.* **2015**, *160*, 15–30. [CrossRef]
11. Bailey, S.W.; Werdell, P.J. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* **2006**, *102*, 12–23. [CrossRef]
12. Giardino, C.; Pepe, M.; Brivio, P.A.; Ghezzi, P.; Zilioli, E. Detecting chlorophyll, Secchi Disk Depth and surface temperature in a sub-alpine lake using Landsat imagery. *Sci. Total Environ.* **2001**, *268*, 19–29. [CrossRef]
13. Lesht, B.M.; Barbiero, R.P.; Warren, G.J. A band-ratio algorithm for retrieving open-lake chlorophyll values from satellite observations of the great lakes. *J. Gt. Lakes Res.* **2013**, *39*, 138–152. [CrossRef]
14. McCullough, I.M.; Loftin, C.S.; Sader, S.A. Landsat imagery reveals declining clarity of Maine's lakes during 1995–2010. *Freshw. Sci.* **2013**, *32*, 741–752. [CrossRef]
15. Johnson, R.; Strutton, P.G.; Wright, S.W.; McMinn, A.; Meiners, K.M. Three improved satellite chlorophyll algorithms for the southern ocean. *J. Geophys. Res. Oceans* **2013**, *118*, 3694–3703. [CrossRef]
16. Olmanson, L.G.; Brezonik, P.L.; Bauer, M.E. Evaluation of medium to low resolution satellite imagery for regional lake water quality assessments. *Water Resour. Res.* **2011**, *47*. [CrossRef]
17. Meuser, J.E.; Baxter, B.K.; Spear, J.R.; Peters, J.W.; Posewitz, M.C.; Boyd, E.S. Contrasting patterns of community assembly in the stratified water column of Great Salt Lake, Utah. *Microb. Ecol.* **2013**, *66*, 268–280. [CrossRef] [PubMed]

18. Klausmeier, C.A.; Litchman, E. Algal games: The vertical distribution of phytoplankton in poorly mixed water columns. *Limnol. Oceanogr.* **2001**, *46*, 1998–2007. [CrossRef]
19. Goel, R.; Myers, L. Evaluation of cyanotoxins in the Farmington Bay, Great Salt Lake, Utah. Available online: http://cdsewer.org/GSLRes/2009_CYANOBACTERIA_PROJECT_REPORT.pdf (accessed on 1 January 2016).
20. Marden, B.; Richards, D. Factors Influencing Cyanobacteria Blooms in Farmington Bay, Great Salt Lake, Utah. Available online: https://www.researchgate.net/profile/David_Richards20/publication/305488678_Factors_Influencing_Cyanobacteria_Blooms_in_Farmington_Bya_Great_Salt_Lake_Utah/links/5790eefe08ae0831552f92ab.pdf (accessed on 12 March 2015).
21. Rushforth, S.R.; Squires, L.E. New records and comprehensive list of the algal taxa of Utah Lake, Utah, USA. *Gt. Basin Nat.* **1985**, *45*, 237–254.
22. Whiting, M.C.; Brotherson, J.D.; Rushforth, S.R. Environmental interaction in summer algal communities of Utah Lake. *Gt. Basin Nat*. **1978**, *38*, 31–41.
23. Wurtsbaugh, W.A.; Marcarelli, A.M.; Boyer, G.L. Eutrophication and Metal Concentrations in Three bays of the Great Salt Lake (USA). Available online: http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1548&context=wats_facpub (accessed on 1 July 2012).
24. Naftz, D.; Angeroth, C.; Kenney, T.; Waddell, B.; Darnall, N.; Silva, S.; Perschon, C.; Whitehead, J. Anthropogenic influences on the input and biogeochemical cycling of nutrients and mercury in Great Salt Lake, Utah, USA. *Appl. Geochem.* **2008**, *23*, 1731–1744. [CrossRef]
25. USGS. Great Salt Lake—Salinity and Water Quality. Available online: https://ut.water.usgs.gov/greatsaltlake/salinity/ (accessed on 5 April 2017).
26. Wurtsbaugh, W.; Marcarelli, A. *Eutrophication in Farmington Bay, Great Salt Lake, Utah 2005 Annual Report*; Central Davis Sewer District: Kaysville, UT, USA, 2006.
27. Utah DEQ. *Utah Lake Report*; Utah Department of Environmental Quality: Salt Lake City, UT, USA, 2006.
28. Hansen, C. *Great Salt Lake Water Quality*; Hydroshare: Cambridge, MA, USA, 2017.
29. Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
30. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
31. Olmanson, L.G.; Brezonik, P.L.; Finlay, J.C.; Bauer, M.E. Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. *Remote Sens. Environ*. **2016**, *185*, 119–128. [CrossRef]

CHAPTER 3

EVALUATING HISTORICAL TRENDS AND INFLUENCES OF METEOROLOGICAL AND SEASONAL CLIMATE CONDITIONS ON INLAND LAKE CHLOROPHYLL A CONCENTRATIONS USING REMOTE SENSING [1]

[1] This chapter has been submitted to *Lake and Reservoir Management* and is currently under review.

Problems associated with excessive amounts of algae (harmful algal blooms or HABs) are widely recognized in lakes and reservoirs throughout the world, in both recreational and drinking water systems (Falconer 1999, Heisler et al. 2008). HABs are often defined as a concentration of algae or phytoplankton that cause harm through production of toxins or disturbing food web dynamics (Anderson et al. 2002). Some HABs produce hepatoxins and neurotoxins that cause mortality in small mammals and birds, and illness (from rashes to respiratory problems) in humans from inhalation or contact (Hunter 1998, Backer et al. 2010). Blooms can harm lake ecology by depleting dissolved oxygen for grazing zooplankton and other aquatic organisms as they decompose and by limiting light penetration for benthic plants (Paerl 1988, Paerl, Fulton et al. 2001, Smith 2003, Paerl and Otten 2013). When competing algae populations and zooplankton are limited, this can promote further growth of the bloom-forming species. Limited light penetration reduces biomass and productivity of oxygen-producing benthic plants, which further restricts available dissolved oxygen (sometimes resulting in fish kills). Limited dissolved oxygen also impacts lake aesthetics and recreation as anoxic conditions encourage the growth of sulfate reducing bacteria, production of hydrogen sulfide gas, and cause an unpleasant rotten-egg odor (Acuña et al. 2017, Chen et al. 2017).

Aesthetic problems and “lake stink” caused by the decomposition of large algal blooms have been a recurring complaint of recreationists and visitors to the Great Salt Lake and nearby lakes in Utah, USA for decades. These lakes have garnered significant concern from the public following several years of large algal blooms, with extremely high (>10,000,000 cells/mL) cyanobacteria cell counts in Utah Lake, and detected cyanotoxins (including nodularin and microcystin) in Utah Lake and Farmington Bay (https://deq.utah.gov/health-advisory-panel/harmful-algal-blooms/harmful-algal-blooms-home). Due to irregular historical records, there is little context for how the magnitude of these events compares to those in the past. However, anecdotal history of poor conditions, these recent extreme HAB events, and increased public concern have sharpened the focus on water quality and management issues in these Utah water bodies. In light of the local concern for HABs, several continuous buoys have been installed in Utah Lake, and there have been increased efforts to provide resources about algal blooms, update the public on current conditions, and participate in ongoing monitoring efforts such as the multi-agency Cyanobacteria Assessment Network (CyAN). In other lakes, technologies such as remote sensing, computer models, and real-time data loggers are being developed to better monitor current conditions and even forecast future conditions (Recknagel et al. 2018). These increased efforts to better understand HABs are echoed by the Environmental Protection Agency (EPA), which has encouraged increased monitoring and study of HABs in the future, using both traditional and nontraditional techniques such as remote sensing (EPA 2017). Additional information is still needed about past conditions to provide science-based evidence that supports monitoring agencies (in determining locations and timing for future monitoring efforts) and policy and management decision making by providing a better historical context for current conditions and understanding of how water quality conditions have changed over a long period of time.

The objective of this study is to use existing field records, readily-available remote sensing imagery, and free or open-source tools to 1) extend the temporal field record of water quality data; 2) use the remotely-sensed record to describe long-term, seasonal, and lake-wide patterns of algal biomass, and 3) explore influences of local meteorological events and climate conditions on algae biomass. We accomplished these objectives by calibrating, applying, and evaluating lake-specific models for the distinct bays in the Great Salt Lake and Utah Lake over a 32-year historical remote sensing record. This study combines knowledge of observed physical processes and patterns in the Great Salt Lake system and data-driven techniques to develop models that complement and inform water quality monitoring efforts that are currently in place for this lake system.

3.1 Study Site

The Great Salt Lake is a remnant of the ancient fresh-water Lake Bonneville, which once stretched across Idaho, Utah, and Nevada. As Lake Bonneville drained over the last 14,000 years, it left the surface water system that includes the Great Salt Lake, the Jordan River, and Utah Lake to the south (Arnow and Stephens 1990) (Figure 3.1). These lakes have changed a great deal since early settlers came to the Salt Lake Valley; many of these changes have been engineered, including construction of pumping systems, drains, diversions, and railroad and automobile causeways. For example, one causeway divides the northern bays (Gunnison, Willard, and Bear River) from the larger, southern portion of the Great Salt Lake, creating vastly different conditions in these two areas. Another causeway further isolates Farmington Bay from the rest of the southern portion. Because these causeways restrict flow and mixing between different sections, the different arms/bays are often considered as distinct lakes. In this paper, we focus on the southern portion of the Great Salt Lake and refer to this portion of the lake as GSL. The combination of the GSL, Farmington Bay, and Utah Lake is referred to as the GSL system.

The GSL is a unique body of water, with an average depth of approximately 4.2 meters. It is characterized as a hyper-saline lake, with salinity ranging from 11-15% in the southern half (USGS 2013). Farmington Bay is approximately 1 meter deep and has between 1-8% salinity. Because these lakes are so shallow, surface area varies considerably with lake level. Currently, GSL has a surface area of about 1735 ha while Farmington Bay has a surface area of about 65 ha. The GSL is the terminal point for several major rivers (Jordan, Weber, and Ogden Rivers), and contains significant land and water-access areas for recreation and bird habitat. Farmington Bay supports a vibrant and diverse ecosystem for millions of migratory birds who feed on the abundant insects and brine shrimp (Cox and Kadlec 1995) and features many popular recreation and camping spots. The GSL plays an important role in both the brine shrimp industry and recreation with one of the few marinas and boat launches on the lake. Heavy use of the GSL through bird-watching, hunting, and other recreation draws hundreds of thousands of visitors each year, providing additional motivation for improving lake health and aesthetics. Additionally, the GSL provides a significant net economic value (between $10.3-58.9 million annually) to publicly owned treatment works by serving as a receiving water body for wastewater discharge. The total contribution of all services and uses of the GSL (industrial, aquaculture, and recreational) to the Utah Gross Domestic Product is estimated at $1.3 billion annually (Bioeconomics, Inc. 2012).

Utah Lake is also very shallow, with an average depth of approximately 2.74

meters and approximate surface area of 340 ha. It has higher salinity levels than most freshwater lakes (around 0.1%) that fluctuate depending on inflows and evaporation (UDEQ 2006). Utah Lake has a number of campgrounds, public access points, and marinas, supporting a variety of recreational activities. Like Farmington Bay, it also supports a large migratory bird population, and is heavily monitored because it is a habitat for the endangered June sucker (Chasmistes liorus) species.

Historically, diatoms have dominated Utah Lake, though many species of chlorophyta (green algae) and cyanobacteria (blue-green algae) have also been documented (Rushforth and Squires 1985). In Farmington Bay, large blooms of cyanobacteria (especially Nodularia) have been observed, whereas the most prevalent algae in the GSL are diatoms and green algae (Wurtsbaugh 2008, Wurtsbaugh et al. 2012).

3.1.1 Previous studies of Algal Blooms in the Great Salt Lake System

As is the case with many large lakes, monitoring water quality at high spatial and temporal resolutions has not been feasible in the GSL system. Traditional sampling techniques are costly and cannot reasonably cover the entire spatial extent of large lakes. In addition to the economic and physical limitations of these sampling techniques, there are also limitations to data sharing (e.g. lags between when data are collected and when they are formatted and published) and differences in sampling methods among the various organizations that monitor water quality in the GSL system. These organizations include the Utah Department of Environmental Quality: Division of Water Quality (UDWQ), the United States Geological Survey (USGS), and the Jordan

River/Farmington Bay Water Quality Council (JRFBWQC). Historically, these organizations have used chlorophyll a (chl- *a*) concentrations to estimate algae biomass at set locations throughout the lake system (Figure 3.2).

In the past, remote sensing of HABs in the GSL system has been limited to single events. One study used Earth Resources Technology Satellite (ERTS-1) and aerial imagery from a single date to describe the distribution of an algal bloom in Utah Lake (Strong 1974); however, there were no estimates of algal biomass. More recently, a study of a bloom in 2016 used remote sensing to calculate the Floating Algae Index and the Normalized Difference Chlorophyll Index (Page et al. 2018). In addition, Utah Lake has unique water colors because of high dissolved solids and this study did not attempt to calibrate a site-specific model; rather, it used models from the literature.

Previous studies of algal bloom drivers in the GSL system have largely focused on field-measured or laboratory-controlled chl- *a* and in-lake characteristics and constituents (e.g. water temperature, nutrient and salinity concentrations) (Wurtsbaugh 2008, Goel and Myers 2009, Larson and Belovsky 2013, Marden et al. 2015). Study of any external factors (e.g. hydrology or meteorology) has been limited to Utah Lake and a single algal bloom event (Page et al. 2018). This study extends previous work by using nontraditional methods (i.e. remote sensing) to estimate chl- *a* and by exploring effects of external factors using a long record of algal bloom conditions.

3.2 Methods

Approaches to remote sensing of chl- *a* can generally be classified as either analytical/semi-analytical (relating apparent optical properties (radiance and reflectance) to inherent optical properties near the surface (backscattering and absorption of light) and then deriving in-water constituents) or empirical/semi-empirical (using statistical relationships between the chl- *a* and apparent optical properties). A review of analytical and empirical approaches to remote sensing of water quality (Matthews 2011) shows that both approaches have been demonstrated for a wide range of inland lake and reservoir systems. While the physical processes modeled by analytical approaches are important to understand in certain contexts, this study uses an empirical approach, due to its relative simplicity and ability to exploit available historical data. Historical records are often missing information about other constituents that affect the inherent optical properties used in analytical models. For example, an analytical model of Utah Lake would require historical information about calcite precipitation that affects water color; however, these data are not readily available. Empirical modeling approaches inherently account for the effects of the conditions that affect spectral features without requiring detailed information on the physical parameters. The following section describes the data and empirical modeling approach we used to estimate chl- *a*, as well as the statistical analyses used to determine influences of climate conditions on algal biomass throughout the GSL system.

3.2.1 Data

3.2.1.1 Field Measurements

We obtained historical field-measured surface chl- *a* concentrations from the UDWQ using the Ambient Water Quality Monitoring System (AWQMS) database (from 1995-2012), from the USGS using the National Water Information System (NWIS) (from

1995-2015), and JRFBWQC (from 1997-2015). Data were limited to measurements during the main growing season (May-September). The number of historical observations for Utah Lake, GSL, and Farmington Bay during these time periods are 325, 1148, and 95, respectively. The time series plots and map of sampling locations (Figure 3.2) for these three organizations highlight the temporal and spatial limitation of the existing records.

Chl- *a* concentrations are the most common index of algal biomass used by local monitoring organizations. For the GSL and Farmington Bay, the data used in calibration had been corrected for the accessory pigment pheophytin. For Utah Lake, the available record mostly contained measurements that were not corrected for pheophytin. In the few available records containing both uncorrected and corrected chl- *a* measurements, the corrected chl- *a* concentrations were generally around 30% lower than the uncorrected concentrations. Therefore, to maintain consistency, we used only the uncorrected chl- *a* concentrations for calibrating the Utah Lake models. The measurements for all of the lakes were taken near the surface, at depths < 1 meter (generally <0.25 meters). For dates where multiple samples were taken, we used the measurement taken nearest the surface. These data are summarized in Table 3.1.

3.2.1.2 Satellite Imagery and Google Earth Engine

Various sensors have been used to obtain remotely sensed reflectance data for estimation of chl- *a*. While the resolution of spectral bands and range of the Landsat sensors are not as well suited for water quality applications as some other sensors (Palmer et al. 2015), Landsat data have been widely used for small-scale chl- *a* mapping applications (Giardino et al. 2001, Duan et al. 2007, Allan et al. 2011) and long-term and regional studies of lake chl- *a* (Brezonik et al. 2005, Duan et al. 2009, Torbick et al. 2013, Allan et al. 2015, Hansen et al. 2015, Ho et al. 2017). This is largely due to the consistent revisit rate, long history, and complete spatial coverage of most large lakes at a relatively high resolution. Landsat-based estimates can provide valuable information about water quality conditions throughout large bodies of water over relatively long periods of time compared to other satellites with more optimal band configurations. Landsat imagery continues to be used in modeling algal blooms in inland lakes because it provides extensive records for observing long-term effects (Ho and Michalak 2017) and because it can supplement limited historical records of other sensors (Ho et al. 2017). The long historical record is especially useful where historical field sampling records are as limited as those for the lakes in the GSL system.

The satellite imagery used for this study was acquired by the NASA Landsat-5 Thematic Mapper (TM), launched in 1984, and Landsat-7 Enhanced Thematic Mapper Plus (ETM+), launched in 1999. These Landsat missions collect imagery covering the major global land surfaces (including all of North America) at a revisit rate of 16 days, meaning the satellite repeats coverage for each scene within the Landsat Worldwide Reference System (WRS), a grid reference system of 233 paths and 248 rows once every 16 days. The GSL system is contained within Path 38, Row 32 and Path 38, Row 31 of the WRS. Spatial resolution for Landsat is 30 meters. There are slight variations in Bands 4 and 7 between the two missions that are assumed to be negligible. This allowed for development of models that can produce estimates with either Landsat-5 or Landsat-7 images.

We obtained imagery data through Google Earth Engine (Gorelick et al. 2017) in order to reduce the burden of data management, storage, and processing. The Earth Engine data archive is a petabyte-scale, cloud-based collection of remotely-sensed imagery data. Earth Engine has been used in a wide variety of land cover/water body classification (Alonso et al. 2016) and flood and surface water mapping (Pekel et al. 2016, Tellman et al. 2016) studies. Its use in water quality applications is limited, but Earth Engine has been demonstrated as a useful means for accessing and processing data for algal bloom monitoring in other lake systems (Ho et al. 2017). We created functions using the Earth Engine Python API to query the online server and return the surface reflectance values for the GSL system as a dataframe that could be used within R statistical software for model development and application. The Landsat surface reflectance products available through Earth Engine are produced by the USGS. These products are generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2006), which converts the top of atmosphere (TOA) reflectance to surface reflectance. The long, continuous record and ready availability of these data products allows for a convenient exploration of the historical record. However, we note that the use of LEDAPS introduces potential errors into the modeling process. The LEDAPS method removes atmospheric contributions to the at-sensor radiance when calculating the remote sensing surface reflectance (R_{rs}); however, the LEDAPS approach can sometimes overcorrect for atmospheric interference, resulting in negative reflectance values over water in the short-wave infrared range (where water surface reflectance is typically very low). This is generally most problematic in deep, clear water bodies that appear very dark and have low signal-to-noise ratio; however, the lakes in the GSL system are generally turbid and shallow. In cases of negative reflectance values, we filtered these data and excluded them from model calibration and application. The Landsat surface reflectance product includes a cloud mask and cloud mask confidence band, which allows filtering of pixels that are obstructed by cloud-cover or haze. We included only cloud-free pixels in the model development and application. We used the cloud mask band (which indicates if the pixel is classified as cloud-free) to filter out pixels with cloud-cover. Additionally, pixels from Landsat 7 data that were affected by scan line corrector error (present in images after May 31, 2003) were not masked from use in model development and application.

We stored spatial information for the sites in the historical field sampling record in a cloud-based Google Fusion Table and referenced this table from the data retrieval function to obtain the latitude and longitude of the 30-meter pixel corresponding to an individual field sample. For each field sample, we downloaded the surface reflectance values (from the pixel corresponding with the field sampling location) for every image within ±10 days of the sample date. If pixels from multiple images were returned for a single field sample, we used the reflectance values from the image nearest the sampling date (unless there was cloud-cover).

3.2.1.3 Climate Data

We obtained historical daily meteorological records matching the record of remotely sensed imagery (1984-2016) from the National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center archives for stations near the Great Salt Lake and Farmington Bay (USW00024127, located at 40° 46' 21"N, -111° 57'

19.01"E) and Utah Lake (USC00428973, located at 40° 22' 0.01"N, -111° 54' 0"W). These stations provide the most complete record for precipitation (mm/day), wind speed (m/s) or total daily movement (km), and maximum/minimum daily temperature (degrees Celsius) near the study area. Additionally, data from Natural Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) sites representing snow water equivalent (SWE) were obtained. The SNOTEL sites used in this study were 820, located at 40° 26' N, -111° 37' W, and 684, located at 40° 46' N, -111° 38' W. Site 820 is assumed to be representative of snowpack for Utah Lake, while site 684 is representative of snowpack for the GSL and Farmington Bay.

3.2.2 Determining Use of Near-Coincident Data

Historical water quality data were collected without respect to the satellite overpass schedule, which means that many of the samples were not coincident (obtained at the same time or even on the same day) as satellite images in the study. Near-coincident data have been used in a number of studies in order to deal with the relatively few coincident matches, though definitions of near-coincident time windows (time between the satellite image acquisition and the field sample used for calibration) vary greatly in the literature. Reported usable time windows for coincident data range from a difference of ±3 hours (Bailey and Werdell 2006), 1 day (Lesht et al. 2013), 7 days (Kloiber et al. 2002, McCullough et al. 2013), to ±10 days (Olmanson et al. 2008). Previous remote sensing studies of nearby lakes in Utah used same-day and ±1 day limits for near-coincident data (Hansen et al. 2015).

We based the definition of near-coincident data for this study on observations of temporal variability in the lake system and availability of near-coincident data (Hansen et al. 2017). These observations exhibited high variability and low autocorrelation for Utah Lake after 2 days, indicating time windows of less than 2 days would be appropriate for Utah Lake. On the other hand, Farmington Bay and GSL exhibited much less variable patterns, indicating more relaxed limits on time windows could be considered for these lakes. To further justify the definition of near-coincident data that would be used for the final models in this study, we fit preliminary stepwise generalized linear models (GLMs) for each time window from 0-10 days. Bands, band ratios, and the Normalized Difference Vegetation Index (NDVI) were considered as potential predictor variables in the stepwise regression. We then compared model performance (as measured with R^2 between the modeled and observed values) for the model with the lowest Akaike Information Criteria (AIC) for each time window. For each lake and season, the R^2 generally decreased as the time window was relaxed (Figure 3.3). The exception to this is the GSL-Summer model, which had a significant increase in the number of observations, a much broader range of chl- *a,* and a better model fit with a time window of ± 2 days. We did not develop any preliminary models for Farmington Bay using time windows less than 2 days due to insufficient data. The final definitions for near coincident data, based on the observational data and the exploration of different time windows, were defined as ± 1 day for Utah Lake, ± 2 days for GSL, and ± 4 days for Farmington Bay.

3.2.3 Model Parameterization and Application

A wide variety of relationships between remotely sensed data and chl- *a* has been suggested in the literature for empirical models in different types of waters (open ocean, coastal, inland lake, shallow, turbid, etc.) (Gurlin et al. 2011, Lesht et al. 2013, Ali and Ortiz 2014). Optimal band selection differs from one body of water to another due to characteristics such as depth (i.e. reflectance of the bottom of the lake) and the presence of other constituents or turbidity that can change the reflectance (i.e. color) of the water (Matthews 2011). However, it is unclear which Landsat-based bands might best be used to model concentrations of chl- *a* in the GSL system as the lakes are unique in terms of water chemistry and appearance. The best bands to use for estimation may also differ depending on the season if there are patterns of succession in algae populations as different algal populations can have different reflective properties (Casterlin and Reynolds 1977, Stadelmann et al. 2001, Hansen et al. 2015). Because each lake in the GSL system differs in physical characteristics and algal speciation, we parameterized and calibrated unique models for each lake. We used seasonal models to exploit the differences in chl- *a* pigmentation in different algae populations, with seasons based on the observed trends in succession of algae populations in the system. There is a strong pattern of algal succession documented in literature, especially for Utah Lake, with large variability in species in the early summer months, while the later months show a reduction in species diversity with the dominating species being largely blue-green algae (Whiting 1978, Rushforth et al. 1981, Rushforth and Squires 1985). To represent these different seasons, we calibrated two models for each lake: one for spring (May-June) and one for summer (July-September).

Our approach to model development is rooted in observations of physical characteristics and processes (by creating unique models for the distinct lakes and seasons and using appropriate near-coincident data) while being data-driven (using statistical techniques to determine model parameterization). We calculated the correlation between surface reflectance data and chl- *a* concentrations using a GLM structure with log-link functions to account for the nonnormal distribution of chl- *a* data (Nelder and Wedderburn 1972). We used a mixed stepwise regression of the GLM based on minimizing the AIC to provide an initial suite of predictor variables. These variables were selected from allTM/ETM+ bands, band ratios, and the NDVI. The AIC is particularly useful for evaluating models with different numbers of parameters, as it penalizes models with more parameters (similar to an adjusted R2 value). We further refined the initial parameters to limit inclusion to the model to only those that were highly significant (p-value<0.05).

We applied the calibrated models to water-masked, cloud-free surface reflectance data from 1984-2016. We computed a modified normalized difference water index (MNDWI), which was developed specifically for distinguishing water bodies in Landsat imagery (Xu 2006), and applied the results to each scene to mask nonwater pixels. This is to account for variation in lake surface area over time. We also excluded any pixels flagged for cloud or cloud shadow. The MNDWI equation uses the mid-infrared band (band 5 in Landsat 5 and 7 imagery), as shown in Equation 3.1.

$$MNDWI = \frac{Green - MIR}{Green + MIR} \tag{3.1}$$

Pixels with positive MNDWI values were considered water, while negative MNDWI is classified as land (Xu 2006). We assumed that any estimated chl- *a* concentrations above 500 µg/L were noise and masked them from the final results.

3.2.4 Statistical Analysis

3.2.4.1 Trends

We calculated long-term trends (change in chl- *a* (µg/L per year) over the resulting historical record using the nonparametric Theil-Sen estimator, which is robust for nonnormally distributed data and outliers. The Theil-Sen slope estimator is calculated as the median of the slopes of the lines crossing all possible pairs of distinct points. Its use in water quality time series analysis has been demonstrated in literature (Esterby 1996), particularly when the time series is irregular or when time series exhibit seasonality (Hirsch et al. 1982). We used the Median-Based Linear Models, or mblm package, in R (Komsta 2013) to calculate the Theil-Sen estimator for each of the lakes over the 32-year constructed record. We calculated annual trends for a number of characteristics describing algal bloom dynamics. These include: mean, 99th percentile (representing the extreme), variability (measured via the standard deviation), and spatial extent (calculated as a percentage of pixels above 50 µg/L, which represents eutrophic conditions, according to the Carlson TSI metric (Carlson 1977)). We also calculated trends for these measures yearly for each month to examine potential seasonality (Hirsch et al. 1982). We calculated trends in the timing of annual maximum chl- *a* values (day of year when the estimated maximum chl- *a* level occurred) using the Theil-Sen estimator.

3.2.4.2 Influence of Weather and Climate Conditions

3.2.4.2.1 Immediate Effects of Local Weather

We explored influences of localized weather conditions on surface chl- *a* concentrations on multiple scales. Meteorological events, including high wind and precipitation events, can contribute to mixing in the shallow lakes of this system. Additionally, high air temperature generally corresponds with warmer water temperature, and may affect the lakes by providing a favorable environment for algae growth. These events that affect surface algae populations and contribute to mixing could have immediate impacts on concentrations of surface chl- *a* over timescales that are relevant to sampling or short-term monitoring. For this study, we defined "events" as days with greater than monthly mean wind speed (m/s) or wind movement (km), nonzero precipitation (measured in mm), or greater than monthly mean daily maximum air temperature (degrees Celsius). To evaluate the immediate effects of these events, we first subset the estimated record of chl- *a* at the historical monitoring locations with one group consisting of observations immediately following (within one day of) an event and the other group consisting of observations that did not follow an event. We then evaluated differences between these subsets in two ways: comparing the mean chl- *a* concentrations for each subset at each location and using the Wilcoxon Rank Sum test (also called the Mann-Whitney test) to determine if the distributions of the two subsets were statistically significant (α=0.05). This test essentially calculates the likelihood that a randomly selected value from one subset will be less than or greater (when using a two-sided test) than a randomly selected value from the other subset. Finally, we performed a simple correlation test between each of the variables and chl- *a* concentrations using the nonparametric Kendall's Tau method. This test measures the strength of the relationship between the weather variable and chl- *a* concentrations, with a null hypothesis that there is no association between the variables.

3.2.4.2.2 Effects of Seasonal Climate Conditions

Analysis of seasonal climate conditions was aimed at evaluating whether the seasonal climate-related conditions that were identified as likely triggers for the 2016 bloom (Page et al. 2018) are supported by evidence in the remotely sensed record. This study, which used modeled climate and hydrology data, suggests that the excessive algal biomass was largely caused by high winter and spring (December – March) precipitation, high spring (March) runoff, and low precipitation and warmer temperatures in early summer months (May-June). Instead of modeled climate and hydrology data, we used the weather observations from NOAA stations (the same as those used in the short-term analysis) and SWE observations during March from the SNOTEL sites as a proxy for runoff (which is largely driven by snowpack, and March is typically when the peak SWE occurs).

To evaluate whether these conditions are consistent with algal bloom conditions in the past, we first calculated precipitation totals, average temperatures, and average SWE for the seasons defined above. The authors of the 2016 bloom study did not specify any thresholds for distinguishing whether the values of these select climate variables were high or low other than comparing values to values in other months in the year leading up to the bloom. This may be problematic, as it mostly describes the general seasonal climate patterns of the study area (higher precipitation in winter and spring, high runoff in the spring, and low precipitation and higher temperatures in early summer). For this reason, we defined thresholds of seasonal means using the 32-year mean to distinguish between high and low values for each of the seasonal climate variables. Then, similar to the evaluation of immediate local weather effects, we subset the 32-year record according to years with the same "favorable" seasonal climate conditions as 2016: high (above historic mean) precipitation in December-March, high SWE (as a proxy for runoff) in March, low (below historic mean) precipitation and high (above historic mean) temperatures in May-June. We then compared mean values of annual maximum lake-wide chl- *a* for the above average and below average subsets, applied the Wilcoxon Rank Sum test to the two subsets, and evaluated general correlation between the seasonal climate variables and the annual maximum chl- *a* using Kendall's Tau.

Additional information about the tools and software packages developed to accomplish the workflow (Figure 3.4) is provided in the supplemental material.

3.3 Results

Final model coefficients and parameters are provided (Table 3.2). Model residuals have no apparent patterns or trends and approximate a normal distribution. This indicates that the predictive abilities of the models are relatively constant over the range of observations. Model goodness of fit was measured via percent bias (PBIAS), which is calculated as follows in Equation 3.2:

$$PBIAS = 100 * \frac{\sum(Mod_i - Obs_i)}{\sum Obs_i} \tag{3.2}$$

where Mod*i* and Obs*i* are corresponding values for each *i*th observation of chl- *a* from the modeled and calibration datasets. The PBIAS for the final models (Table 3.2) ranges from -15.9 to17%. Positive values indicate overestimation bias, whereas negative values indicate model underestimation bias. The R^2 between the modeled and observed chl- *a* for each of the models ranges from 0.79 to 0.99 (Table 3.2 and Figure 3.5). Mean Absolute Error (MAE) between modeled and observed values was also calculated and is generally low (1-13.4 µg/L), with the exception of the Spring Farmington Bay model, which has a MAE of 38.5 µg/L, where there were relatively few data available for calibration and the magnitudes and range of concentrations were much greater than other lakes and seasons.

These models produce a remotely-sensed record that supplements and enhances the field sampling record by greatly increasing the extent and the number of chl- *a* estimates (Figure 3.6). The remotely-sensed estimates extend the field sampling records more than a decade earlier than the field record, and in many cases, provide more complete seasonal records than the field sampling record. There were approximately 7 times more remotely sensed estimates than field samples for Utah Lake and GSL, and 34 times more for Farmington Bay (which had the most severely limited field sampling record). The increase in the number of data available for analysis reduces the uncertainty in the results of statistical analyses of trends and relationships. The models also increase the spatial information available for the lake system with long records of lake-wide statistics, such as averages and extremes (Figure 3.7).

3.3.1 Trends in the Constructed Historical Record

Over the 32-year remotely-sensed record, the lake system shows only minimal differences for average values and spatial extent of blooms. However, there are significant increasing trends in the extremes (99th percentile) for all lakes (1.8-4.9 µg/L per year) and variability (0.2-1.2 µg/L per year) (Table 3.3). Long-term trends in the timing of blooms (i.e. day of year when average chl- *a* and extreme chl- *a* concentrations peaked) exhibit a statistically significant shift to earlier in the growing season. This ranged from 1.2-2.5 days earlier per year for peaks in average concentrations and 0.5-2.5 days earlier per year for peaks in extreme values.

Seasonally, the long-term trends in the lake system are more nuanced (Figure 3.8). The data show trends of increasing averages and extremes in the system in May and June, while July through September generally exhibit decreasing trends (or in some cases, smaller positive trends compared to those observed in the earlier months). The earlier months also exhibit trends of increasing variability and in the case of the GSL, increasing spatial extent of blooms. The pattern of increasing trends in algal bloom dynamics in earlier months, along with the observed shifts in the timing of peak average and extreme concentrations, indicates that poor conditions in the lake system are generally occurring earlier in the season.

3.3.2 Short-term Influences of Weather

Comparisons between average chl- *a* concentrations following precipitation, above average wind, and above average temperature events identified locations where algae biomass may be especially sensitive to the influence of local weather. The effects of each of these variables varies within the lake system, and in some cases, within each lake (Table 3.4).

3.3.2.1 Precipitation

Precipitation events had pronounced effects on sites in Utah Lake and Farmington Bay near stream inflows. In Utah Lake just outside of Provo Bay and at the southern end of the lake that receives inflows from springs and surface runoff from undeveloped/agricultural lands, mean chl- *a* concentrations were greater following precipitation events (10 µg/L and 15 µg/L, respectively). The site inside of Provo Bay, however, had lower mean chl- *a* concentrations (20 µg/L) following precipitation events. The southernmost site in Farmington Bay, which is nearest the inflow from the Jordan River, also had lower mean chl- *a* concentrations (20 µg/L) following precipitation events. Other locations in the lakes showed minimal differences in mean chl- *a* concentrations. Differences in distributions of the precipitation subsets were generally not statistically significant with the exception of the Provo Bay locations. There were no significant correlations between any of the sites and precipitation amounts in any of the lakes. However, differences in mean chl- *a* concentrations for select sites in Utah Lake suggest that additional monitoring at these locations may be warranted after rainfall events. Because of the proximities to streams and surface runoff into the lakes, these locations may be responding to the complex effects of concentrated nutrients that are flushed into the streams during storm events, increased freshwater inflow, and movement of algae.

3.3.2.2 High Wind

For estimates of chl- *a* following high wind events (above mean 24-hour wind movement or mean 24-hour wind speed), effects were also highly localized in Utah Lake. For Utah Lake, average concentrations were substantially lower (>30 µg/L) following high wind events than on dates with lower wind movement at the site within Provo Bay on the east side of the lake, while the station across the lake demonstrated the opposite behavior with small increases (<5 µg/L). This was expected, as the prevailing wind direction is east-west, and the difference in average concentration may at least partially be explained by wind transport of the algae. Minimal differences were seen at other sites in Utah Lake and in the GSL, while small increases (on average 6 µg/L) were seen for most locations in Farmington Bay following high wind events. This suggests that wind may have influence on sampling in Farmington Bay and parts of Utah Lake, but has little impact on sampling in the GSL. Differences in distributions at the sites in Provo Bay and on the west side of Utah Lake were statistically significant (though differences at other sites in Utah Lake and the other lakes were not). Small, but statistically significant correlations between wind movement and chl- *a* were seen for the site within Provo Bay (-0.15) and on the west side of Utah Lake (0.11), offering further evidence of east-west movement of surface algae during wind events. Very small but significant positive correlations between wind speed and chl- *a* (0.07-0.09), were observed for several sites in the GSL and Farmington Bay, but most sites did not have any significant correlation.

3.3.2.3 High Air Temperature

Above average temperatures were typically followed by higher mean chl- *a* concentrations for Utah Lake sites (on average 9 µg/L higher) and Farmington Bay (on average 13 µg/L higher). The difference in the site within Provo Bay was the most pronounced, where the mean of samples taken following days of above average temperature was much higher compared to the mean of samples following days of below average temperature (about 55 µg/L). In contrast with these lakes, the GSL sites consistently had lower mean chl- *a* concentrations (on average 20 µg/L lower) when samples followed days of above average temperature. Differences in distributions of the subsets for most sites in Utah Lake and Farmington Bay were not statistically significant, while the differences in distributions of the subsets for Utah Lake were. Statistically significant correlations between air temperature and chl- *a* at the Provo Bay sites were positive (from 0.1 to 0.16), while several other sites had significant negative correlations (-0.12 to -0.1). The sites at the northern and southern ends of Farmington Bay had small, but statistically significant positive correlations (0.08 to 0.09). Every site in the GSL had statistically significant negative correlations (ranging between -0.27 and -0.07). These observations suggest that higher air temperatures do not lead to favorable for the algae populations in the GSL sites, while they may encourage growth for the algae populations in some areas of Utah Lake and Farmington Bay.

3.3.3 Influence of Seasonal Climate Conditions

The historical record offered supporting evidence for some of the seasonal climate conditions suggested by Page et al. to have caused the 2016 bloom in Utah Lake, though not all relationships were consistent, and there were some differing patterns observed across the three lakes (Figure 3.9). For all lakes, the mean of annual maximum chl- *a* was higher with above average winter precipitation, and there was also a greater mean with above average SWE for Utah Lake. However, there were not major differences for Utah Lake in the subsets for summer precipitation or summer temperatures. Differences in mean values for the GSL subsets were greatest for winter precipitation, and summer temperature, with a higher mean for the below average summer temperatures subset. As with the evaluation of short-term weather effects, this suggests that chl- *a* concentrations in the GSL are greater under cooler summer temperature conditions. The only statistically significant differences in distributions of above and below average subsets (according to the Wilcoxon Rank Sum test) were for winter precipitation and SWE in Utah Lake. These were also the only seasonal climate variables that had any significant correlations (according to the Kendall's Tau test). For Utah Lake, there were positive correlations of 0.34 and 0.37 for winter precipitation and SWE, respectively. Despite the lack of statistically significant differences and correlations between summer temperatures, it is worth noting that the long-term trend of increasing chl- *a* extremes coincide with an increasing trend of summer temperatures that is not statistically significant (with a positive Theil-Sen estimate of 0.02 degrees C/year for both weather stations used in this study). This may indicate a less direct relationship than was suggested by the Page et al. study; as temperatures are generally increase, this extends the growing season earlier (which makes sense with the observed earlier shifts in timing of chl- *a* maxima).

3.4 Discussion

Analysis of long-term trends indicate there has been a shift over the past several decades in the timing of algal blooms, with positive trends in the earlier months for several algal bloom dynamics and trends for peak conditions shifting earlier in the growing season. There are also positive trends in the annual extremes and variability over the last three decades. These increasing trends are occurring in a lake system where local summer temperature also increased. The shifting in timing and positive trends in extremes could be especially problematic in the face of changing climate conditions, such as differences in precipitation patterns and warming temperatures, which could create favorable conditions for extending the season of algal blooms. Future monitoring and management efforts may need to adjust for the seasonal shifts in order to plan for the emergence of blooms and poor conditions.

Localized sensitivity to short-term meteorological events provides important context for evaluating field sampling results (whether the sampled concentrations might be temporarily higher or lower because of a short-term meteorological condition). This could also aid in predicting likely movement of algae and areas that are susceptible to increases following storms, high winds, or periods of above average temperatures. Variability in responses to short-term events also highlights the complexity of interactions between these influencing factors and water quality measures. Wind and mixing of shallow waters may have complex, and even conflicting, interactions that influence algae growth. In many shallow lakes, resuspension of sediments can be caused by wind, which can sometimes lead to more available phosphorus. However, as suspended solids increase, turbidity increases, limiting light penetration needed for growth (Søndergaard et al. 2003). Knowledge of where blooms are likely to concentrate, spread, or move due to wind force/mixing could be helpful in tracking blooms, targeting sampling efforts, and making decisions dealing with closure or warnings for public recreational areas. Additionally, patterns of algal bloom movement may help inform hydrodynamic modeling efforts by providing insight into the sensitivity of algal blooms to certain factors like wind speed and surface mixing.

The long-term remotely-sensed record supports this suggestion as high winter/spring precipitation and average/below average summer precipitation coincided with the major increases in extremes (in 1993 and 2005) for Utah Lake. Farmington Bay and GSL appear to be less responsive to seasonal precipitation (and subsequent surface runoff) than Utah Lake, showing instead more general increases in chl- *a* extremes coincident with higher summer temperature. These differences underscore the complexity of the lake system and the factors that influence algal bloom dynamics.

The trends for the estimated record (limited to estimates at historical monitoring locations) differ significantly from those observed in the field sampling record. This is due to the large differences in record length and frequency. For example, trends in the field sampling record indicate a much larger increasing trend in Utah Lake averages (3.1 µg/L per year) and a decreasing trend in the GSL (-1.3 µg/L per year). Inspection of the estimated and field sampling records reveals the cause of the major difference in trends is an increase in the frequency of sampling over the field sampling record of Utah Lake, and a substantial decrease in the frequency of sampling for GSL. Likewise, trends in timing of annual maxima differ from those in the field sampling record. The remotely sensed record indicates an earlier shift in timing for all of the lakes, while the field sampling record indicates a trend of later peak conditions for Utah Lake. According to the field sampling record, the timing of the highest average chl- *a* concentrations is occurring 2.7 days later each year, while the remotely sensed record indicates that the timing has shifted 1.2 days earlier each year. The differences in trends for the estimated and field-sampled records suggest that the irregular and inconsistent nature of the field sampling record could lead to incorrect conclusions about water quality behavior in the lake system.

3.5 Conclusions

This study identified several long-term changes in the water quality of the GSL system over the past several decades: shifts in when chl- *a* concentrations are the greatest, increasing extremes, and increasing variability of chl- *a* throughout the lakes. Lake managers must be able to anticipate and prepare for these changes by considering how future climate conditions could alter behavior and health of the lake system. Locations that are particularly sensitive to precipitation, high winds, and high temperatures may warrant additional consideration as the lakes are monitored in the future, and future sensor placement or timing/frequency of monitoring may need to be adjusted to account for this sensitivity.

The results and the patterns observed in the GSL system demonstrate the utility in having an enhanced historical record, which provides more frequent and consistent observations as well as greater spatial coverage than historical field sampling. This enhanced record allows for exploration of long-term trends, spatial patterns, and relationships to local climate conditions that are not feasible with limited field sampling-based records. The trends and connections to short-term climate events and longer-term seasonal climate conditions should be used to guide local environmental and regulatory agencies as they focus resources and plan future monitoring efforts. In addition to the influences of the climate conditions explored in this study, other factors, such as nutrient loading from point and nonpoint sources in the surrounding area, should be studied to examine influence on the algal dynamics in the lake system. One of the main obstacles in doing so will be obtaining a long-term record of these contributions of nutrients.

While the remotely sensed record provides a number of benefits, it is important to note the limitations of this study. These include sub-optimal band configuration, surface reflectance products, and revisit rate of historical Landsat imagery. The 16-day revisit rate could potentially miss entire bloom events, which could limit the ability to determine the true trends and timing of peak chl- *a* concentrations. However, the remotely sensed record based on Landsat imagery still provides a much more complete view of long-term conditions than field records. Additionally, the failure of the Scan Line Correction in Landsat 7 images resulted in the loss of some data (pixels masked by the scan lines were not available for use in calibration or application). Another limitation to the model development were the relatively few historical data available for calibration (especially in Utah Lake and Farmington Bay). As a consequence of the limited calibration data and limitations of Landsat data, there were relatively large errors for the Farmington Bay spring model. These limitations may be overcome and the models may be improved as sustained records of better remote sensing products (such as newer satellites including Landsat 8 and Sentinel-2) become available and with increased monitoring that is coordinated with imagery acquisition. Several studies have demonstrated the use of these satellites for mapping of chl- *a* (Manzo et al. 2015, Toming et al. 2016). These instruments have great potential to build on long-term studies using older instruments and can provide data for recent history and ongoing monitoring applications. Another limitation of this study is in the use seasonal approximations of algal succession and chl- *a* as a measure of biomass. Additional information about seasonal algal speciation (which is not currently available for the historical record) would improve the definition of seasonal models. Frequent collection of other measures, such as phycocyanin concentrations, would provide valuable information about whether the algal blooms are potentially toxic. As the improved remote sensing data and field data become available, the workflow and tools presented in this study for obtaining remotely sensed data, calibrating, and applying models can be adapted and applied for future analysis and monitoring.

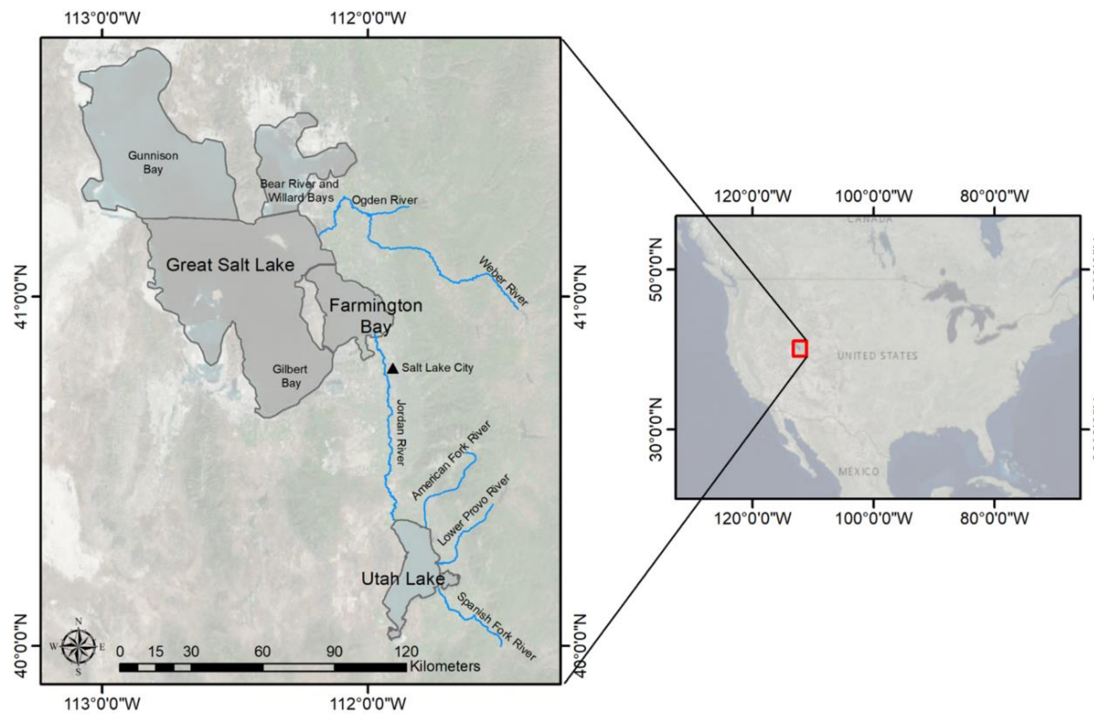

Figure 3.1 Great Salt Lake System, with Utah Lake Draining into the Great Salt Lake via the Jordan River. The southern arm of the lake is further divided by a causeway separating Farmington Bay from the rest of the Great Salt Lake.

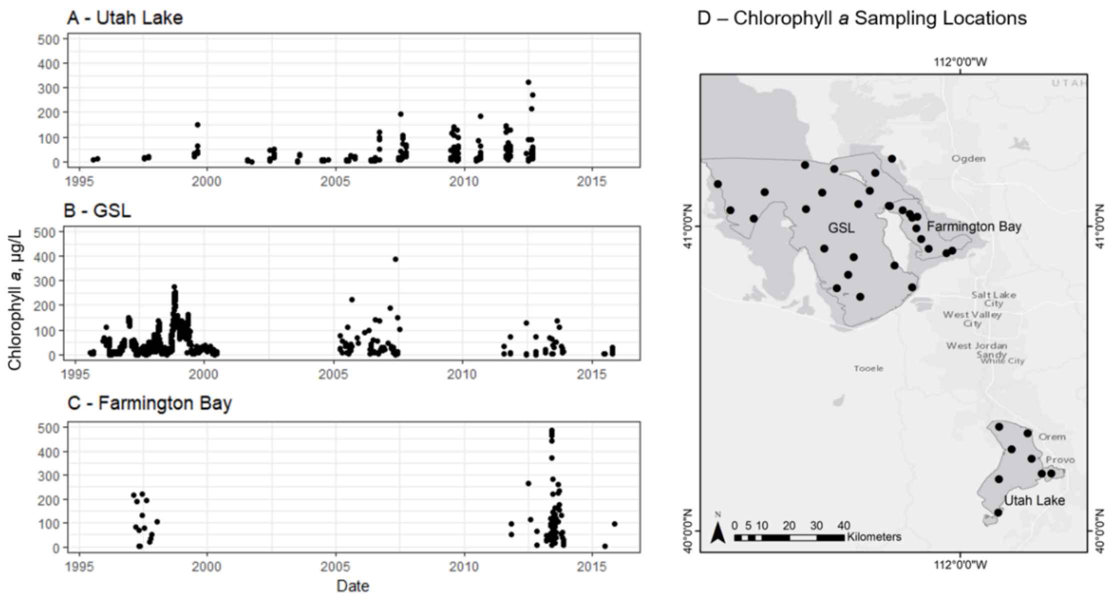

Figure 3.2 Field Record of Chl *a* for A) Utah Lake, B) Great Salt Lake, and C) Farmington Bay between 1995-2015. Data are reported for three organizations (Utah Division of Water Quality (UDWQ), United States Geological Survey (USGS), and Jordan River – Farmington Bay Water Quality Council (JRFBWQC)), for locations shown on D) the map of the study area.

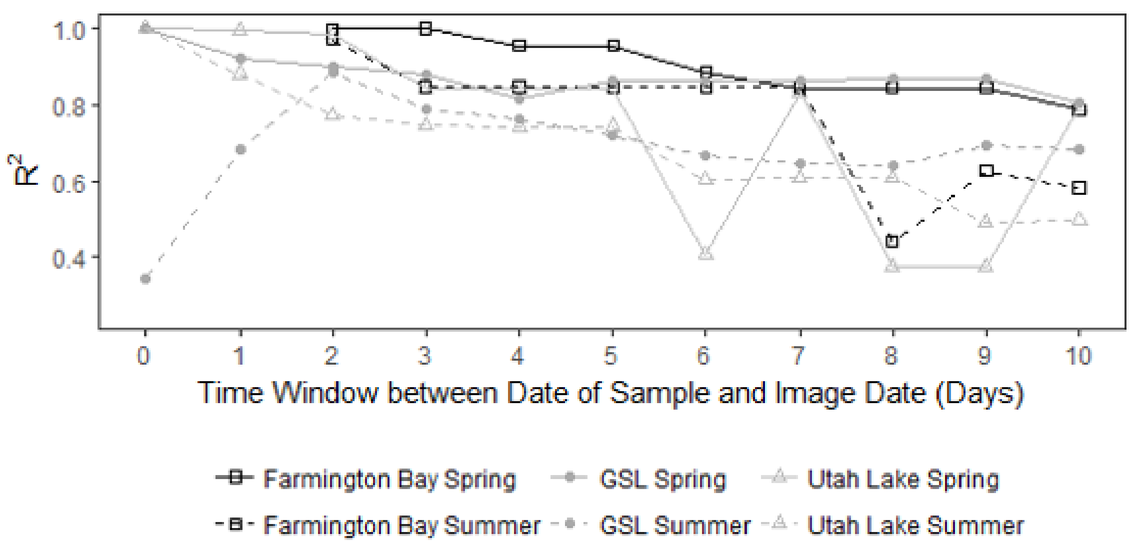

Figure 3.3 Model fit (R^2) between the Modeled and Observed Values for the Initial Stepwise Regression Models by Time Window

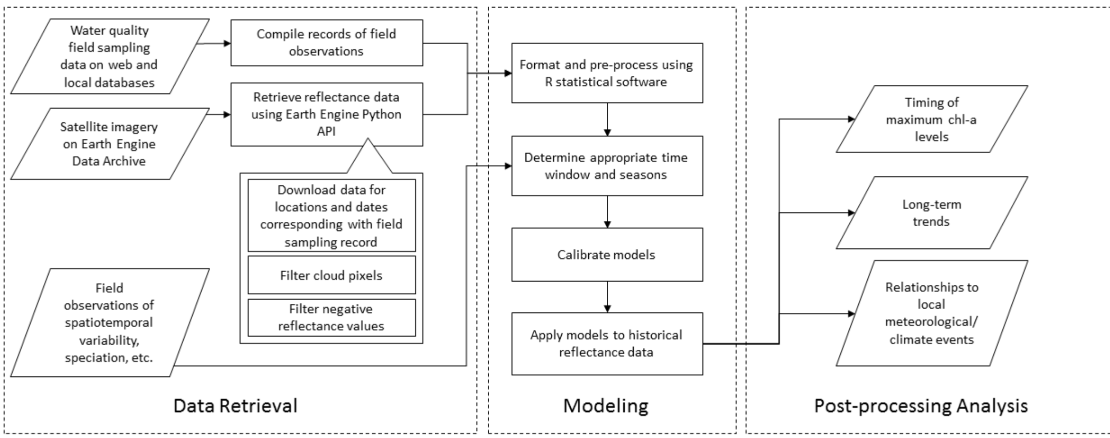

Figure 3.4 Data Inputs and Workflow for Data Collection, Processing, and Model Development, and Postprocessing Analysis

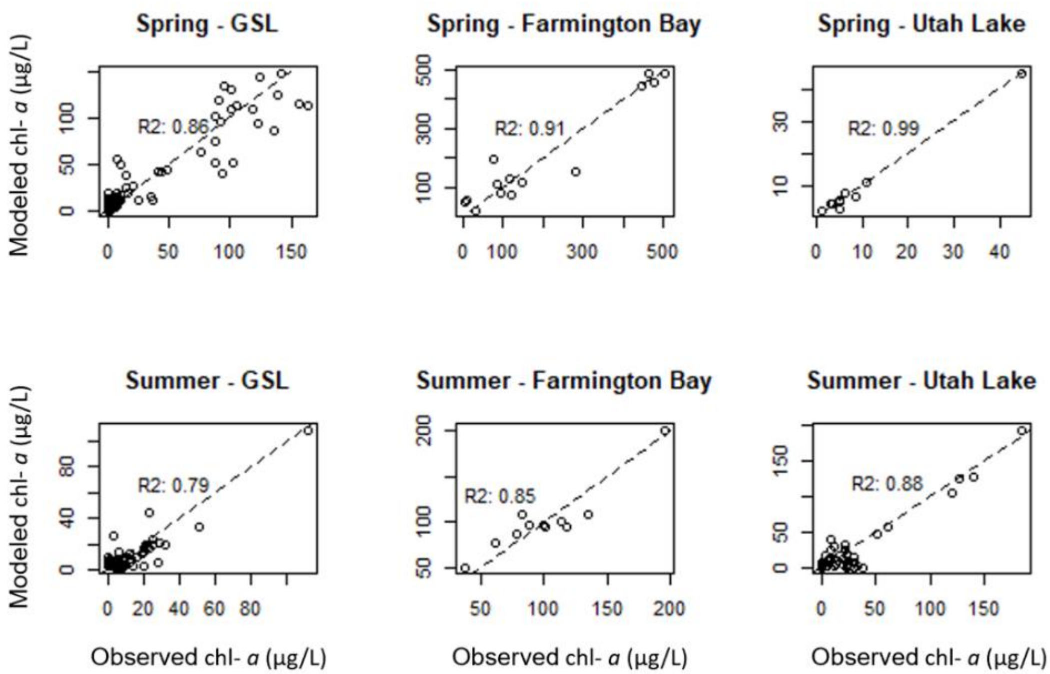

Figure 3.5 Modeled Versus Observed Chl- *a*, with R^2 between Modeled and Observed Values

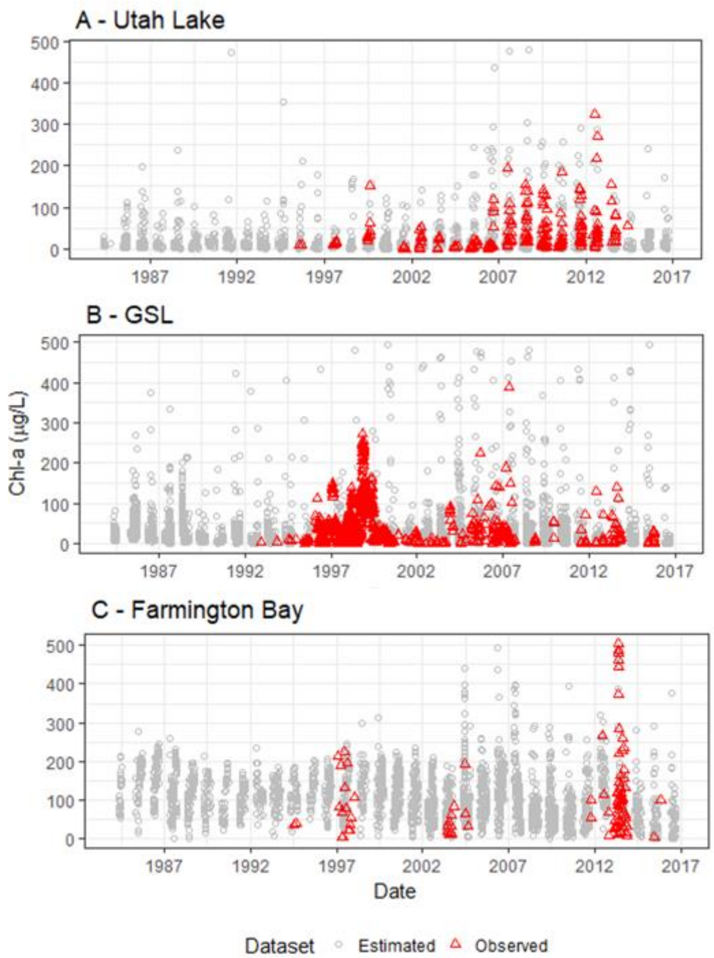

Figure 3.6 Historical Record of Chl- *a* Concentrations for the Sampling Locations in the Great Salt Lake System A) Utah Lake, B) Great Salt Lake, and C) Farmington Bay. The observed dataset is from field samples and the estimated is from remotely sensed imagery.

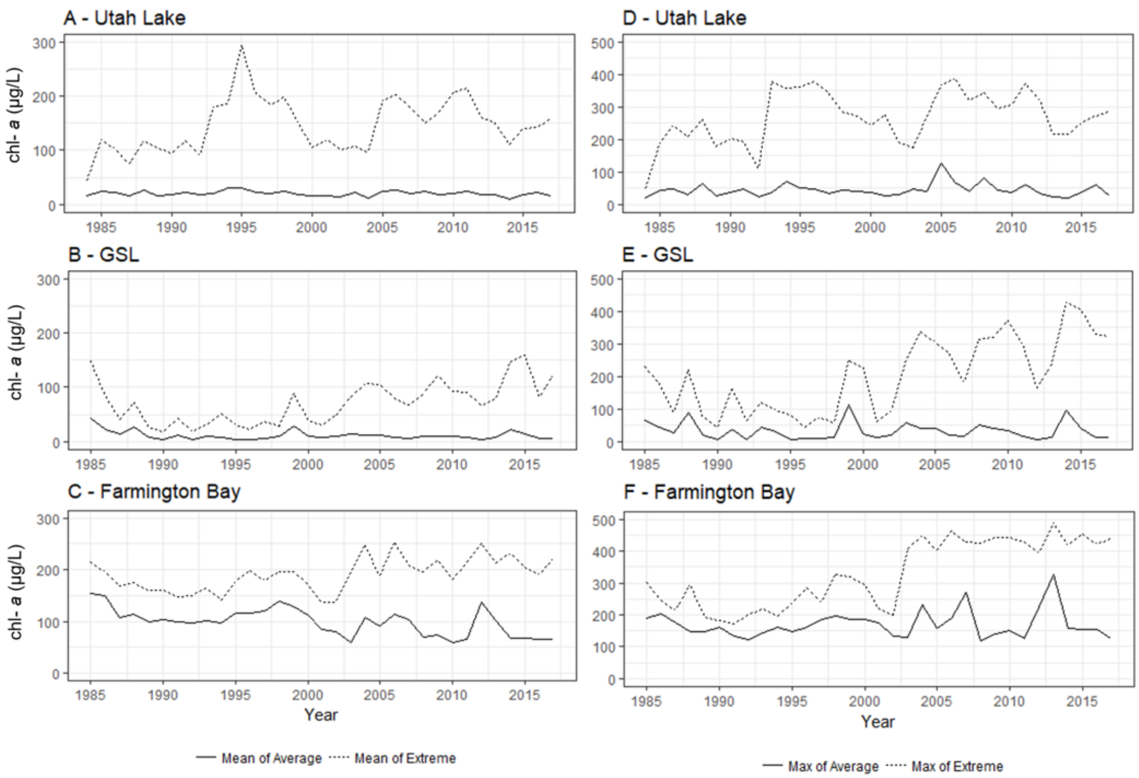

Figure 3.7 Historical Record of Estimated Average and Extreme (99th percentile) Chl- *a* Concentrations. Annual mean of the lake-wide average and extreme are shown in A-C while the annual maximum of the lake-wide average and extreme are shown in D-F. While there is little trend in the measures of average concentrations, increases in measures of the extreme concentrations can be seen for all lakes, particularly after 2002.

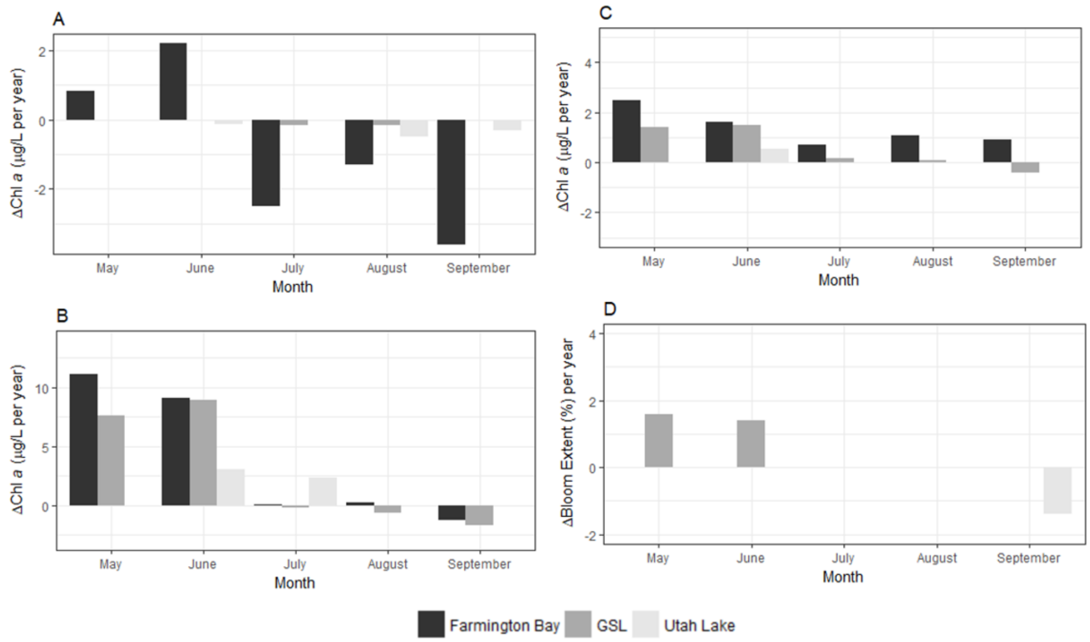

Figure 3.8 Magnitudes of Trends for Lake-Wide A) Average (mean) chl- *a*, B) Extreme (99th percentile) chl- *a*, C) Variability (standard deviation) of chl- *a*, and D) Bloom extent (as a percent of lake >50 µg/L). Only statistically significant trends are shown. Note increasing trends in early summer months for average and extreme concentrations, increasing trends in variability for the whole growing season, and increasing trends for bloom extent in the early summer months in the GSL.

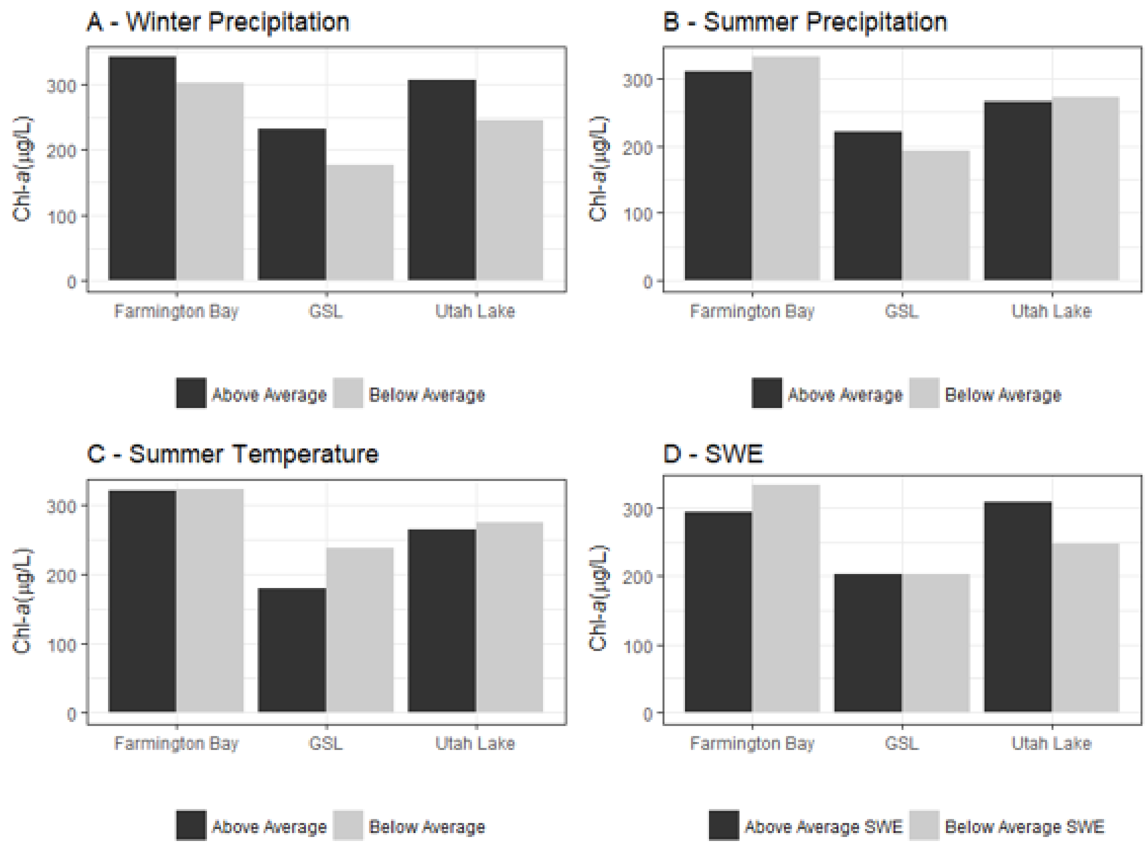

Figure 3.9 Historical Record for Weather Stations East of Utah Lake: A) Seasonal precipitation totals, B) Seasonal average daily minimum temperature, and east of GSL and Farmington Bay: C) Seasonal precipitation totals, D) Seasonal average daily minimum temperature

Table 3.1 Summary Data Used for Model Development, Application, and Analysis

Data Source	**Data Type**	**Period of Record**	**Resolution or Units**
UDEQ or UDWQ	Surface Chlorophyll-a	1992-2009	µg/L
USGS	Surface Chlorophyll-a	1995-2015	µg/L
Jordan River/Farmington Bay Water Quality Council	Surface Chlorophyll-a	2013	µg/L
Landsat-5 TM	Surface Reflectance	1984-2007	30-meter pixels, 16-day satellite overpass
Landsat-7 ETM+	Surface Reflectance	1999-2016	30-meter pixels, 16-day satellite overpass
National Climate Data Center	Precipitation, wind speed, total daily movement, daily maximum air temperature	1984-2016	(mm/day), (m/s), (km), (degrees Celsius)
Natural Resources Conservation Service	Snow water equivalent (SWE)	1984-2016	(in/day)

Table 3.2 Summary of Model Characteristics and Performance. Performance is measured via R^2, Mean Absolute Error (MAE), and Percent Bias (PBIAS) between modeled and observed chl- *a* concentrations.

Lake	**Season**	**Number of Near Coincident Matches**	**Time Window (± days)**	**Range of chl- *a* in calibration data (µg/L)**	**Modeled vs. Observed R^2**	**MAE (µg/L)**	**PBIAS**	**Model**
Utah Lake	May-June	10	1	1.2-45	0.99	0.9	-0.3%	Chl- *a* = exp(-14.23 + 9.33*Green/Blue + 0.003*Blue - 0.004*SWIR1)
	July-Sept	46	1	0.2-185.5	0.88	10.6	-15.9%	Chl- *a* = exp(7.33 - 0.004*Blue - 0.05*Green/SWIR2 + 0.01*Red/SWIR1)
Great Salt Lake	May-June	90	2	0.3-164.2	0.86	10.9	17.0%	Chl- *a* = exp(3.51 + 0.09*Green/Blue - 0.01*Blue + 0.009*Red)
	July-Sept	69	2	0.045-112	0.79	5.2	-3.0%	Chl- *a* = exp(-0.36 + 6.12*Red/Blue - 0.006*Red)
Farmington Bay	May-June	14	4	4-505.9	0.87	38.5	1.4%	Chl- *a* = exp(6.53 - 1.02*Red/NIR - 0.009*SWIR2 + 0.004*SWIR1)
	July-Sept	11	4	37.38-196	0.86	13.4	0.2%	Chl- *a* = exp(6.11 - 0.002*Green)

Table 3.3 Summary of Statistically Significant Long-term Trends (change in µg/L per year), Calculated Using the Theil-Sen Estimator

Lake	**Average Chl-*a* (µg/L/year)**	**Extreme Chl-*a* (µg/L/year)**	**Standard Deviation of Chl-*a* (µg/L/year)**
Utah Lake	-0.25	1.8	0.2
Great Salt Lake	-	2.8	0.55
Farmington Bay	-	4.9	1.2

Table 3.4 Summary of Effects of Meteorological Conditions on Mean Chl- *a* Concentrations throughout the GSL System

Lake	Precipitation Event	High Wind	High Temperature
Utah Lake	Increases at location just outside of Provo Bay (10 µg/L) and at southern end of lake (15 µg/L), Decreases inside of Provo Bay (20 µg/L); Minimal response to precipitation events at other locations	Large decrease after high wind events for site within Provo Bay (>30 µg/L) and small increase at site on west side of lake (<5 µg/L); Minimal response at other locations	Increases for most locations (on average 9 µg/L), large increase at site inside of Provo Bay (55 µg/L)
GSL	Minimal response to precipitation events	Minimal response to wind events	Decreases for all sites (on average 20 µg/L)
Farmington Bay	Decreases at southern end near Jordan River inflow (20 µg/L); Minimal response to precipitation events at other locations	Increases following high wind events for most locations (on average 6 µg/L)	Increases for most sites (on average 13 µg/L)

3.6 References

Acuña, W. C., Gonzalez, C. J., and Aqueveque, V. G. (2017). "La Chimba, Antofagasta, Chile – Oxygen depletion and hydrogen sulfide gas mitigation due to harmful algal blooms." *Harmful Algal Blooms (HABs) and Desalination: A Guide to Impacts, Monitoring, and Management*, edited by D. M. Anderson, S.F.E. Boerlage, and M. B. Dixon, Intergovernmental Oceanographic Commission, Paris, France.

Ali K., Witter, D., and Ortiz, J. (2014). "Application of empirical and semi-analytical algorithms to MERIS data for estimating chlorophyll a in Case 2 waters of Lake Erie." *Environmental Earth Sciences*, 71(9), 4209-4220.

Allan, M. G., Hamilton, D. P., Hicks, B. J., and Brabyn, L. (2011). "Landsat remote sensing of chlorophyll a concentrations in central North Island lakes of New Zealand." *International Journal of Remote Sensing*, 32(7), 2037-2055.

Allan, M. G., Hamilton, D. P., Hicks, B. J., and Brabyn, L. (2015). “Empirical and semi-analytical chlorophyll a algorithms for multi-temporal monitoring of New Zealand lakes using Landsat.” *Environmental Monitoring and Assessment*, 187(6), 364.

Alonso, A., Muñoz-Carpena, R., Kennedy, R. E., and Murcia, C. (2016). "Wetland landscape spatio-temporal degradation dynamics using the new Google Earth Engine cloud-based platform: Opportunities for non-specialists in remote sensing." *Transactions of the ASABE*, 59(5), 1331-1342.

Anderson, D. M., Glibert, P. M., and Burkholder, J. M. (2002). "Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences." *Estuaries*, 25(4), 704-726.

Arnow, T., and Stephens, D. W. (1990). Hydrologic characteristics of the Great Salt Lake, Utah, 1847-1986. *US Geological Survey Water-Supply Paper*.

Backer, L. C., McNeel, S. V., Barber, T., Kirkpatrick, B., Williams, C., Irvin, M., Zhou, Y., Johnson, T. B., Nierenberg, K., and Aubel, M. (2010). "Recreational exposure to microcystins during algal blooms in two California lakes." *Toxicon*, 55(5), 909-921.

Bailey, S. W., and Werdell, P. J. (2006). "A multi-sensor approach for the on-orbit validation of ocean color satellite data products." *Remote Sensing of Environment*, 102(1), 12-23.

Bioeconomics, Inc. (2012). Economic Significance of the Great Salt Lake to the State of Utah. *Report for the Great Salt Lake Advisory Council*, <http://www.fogsl.org/issuesforum/2012/wp-content/uploads/2012/05/Myers_GSLAdvisoryCouncil_Economics.pdf> (June 19, 2018)

Brezonik, P., Menken, K., and Bauer, M. (2005). "Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM)." *Lake and Reservoir Management*, 21(4), 373-382.

Carlson, R. E. (1977). "A trophic state index for lakes." *Limnology and Oceanography*, 22(2), 361-369.

Casterlin, M. E., and Reynolds, W. W. (1977). "Seasonal algal succession and cultural eutrophication in a north temperate lake." *Hydrobiologia*, 54(2), 99-108.

Chen, L., Delatolla, R., D'Aoust, P. M., Wang, R., Pick, F., Poulain, A., and Rennie, C. D. (2017). "Hypoxic conditions in stormwater retention ponds: potential for hydrogen sulfide emission." *Environmental Technology*, 1-12.

Cox, R. R., and Kadlec, J.A. (1995). "Dynamics of potential waterfowl foods in Great Salt Lake marshes during summer." *Wetlands*, 15(1), 1-8.

Duan, H., Zhang, Y., Zhang, B., Song, K., and Wang, Z. (2007). "Assessment of chlorophyll-a concentration and trophic state for Lake Chagan using Landsat TM and field spectral data." *Environmental Monitoring and Assessment*, 129(1-3), 295-308.

Duan, H., Ma, R., Xu, X., Kong, F., Zhang, S., Kong, W., Hao, J., and Shang, L. (2009). "Two-decade reconstruction of algal blooms in China's Lake Taihu." *Environmental Science & Technology*, 43(10), 3522-3528.

Esterby, S. R. (1996). "Review of methods for the detection and estimation of trends with emphasis on water quality applications." *Hydrological Processes*, 10(2), 127-149.

Falconer, I. R. (1999). "An overview of problems caused by toxic blue–green algae (cyanobacteria) in drinking and recreational water." *Environmental Toxicology*, 14(1), 5-12.

Giardino, C., Pepe, M., Brivio, P. A., Ghezzi, P., and Zilioli, E. (2001). "Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery." *Science of the Total Environment*, 268(1-3), 19-29.

Goel, R., and Myers, L. (2009). "Evaluation of cyanotoxins in the Farmington Bay, Great Salt Lake, Utah." Project Report.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote Sensing of Environment*, 202, 18-27.

Gurlin, D., Gitelson, A. A., and Moses, W. J. (2011). "Remote estimation of chl-a concentration in turbid productive waters—Return to a simple two-band NIR-red model?" *Remote Sensing of Environment*, 115(12), 3479-3490.

Hansen, C. H., Williams, G. P., Adjei, Z., Barlow, A., Nelson, E. J., and Miller, A. W. (2015). "Reservoir water quality monitoring using remote sensing with seasonal models: case study of five central-Utah reservoirs." *Lake and Reservoir Management*, 31(3), 225-240.

Hansen, C. H., Burian, S. J., Dennison, P. E., and Williams, G. P. (2017). "Spatiotemporal Variability of Lake Water Quality in the Context of Remote Sensing Models." *Remote Sensing*, 9(5), 409.

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., Dortch, Q., Gobler, C. J., Heil, C. A., and Humphries, E. (2008). "Eutrophication and harmful algal blooms: a scientific consensus." *Harmful Algae,* 8(1), 3-13.

Hirsch, R. M., Slack, J. R., and Smith, R. A. (1982). "Techniques of trend analysis for monthly water quality data." *Water Resources Research,* 18(1), 107-121.

Ho, J. C., and Michalak, A. M. (2017). "Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading." *Journal of Great Lakes Research,* 43(3), 221-228.

Ho, J. C., Stumpf, R. P., Bridgeman, T. B., and Michalak, A. M. (2017). "Using Landsat to extend the historical record of lacustrine phytoplankton blooms: A Lake Erie case study." *Remote Sensing of Environment*, 191, 273-285.

Hunter, P. (1998). "Cyanobacterial toxins and human health." *Journal of Applied Microbiology,* 84(1), 35-40.

Kloiber, S. M., Brezonik, P. L., Olmanson, L. G., and Bauer, M. E. (2002). "A procedure for regional lake water clarity assessment using Landsat multispectral data." *Remote Sensing of Environment,* 82(1), 38-47.

Komsta, L. (2013). mblm: Median-Based Linear Models. R package version 0.12. <https://CRAN.R-project.org/package=mblm> (June 19, 2018)

Larson, C. A., and Belovsky, G. E. (2013). "Salinity and nutrients influence species richness and evenness of phytoplankton communities in microcosm experiments from Great Salt Lake, Utah, USA." *Journal of Plankton Research,* 35(5), 1154-1166.

Lesht, B. M., Barbiero, R. P., and Warren, G. J. (2013). "A band-ratio algorithm for retrieving open-lake chlorophyll values from satellite observations of the Great Lakes." *Journal of Great Lakes Research,* 39(1), 138-152.

Manzo, C., Bresciani, M., Giardino, C., Braga, F., and Bassani, C. (2015). “Sensitivity analysis of a bio-optical model for Italian lakes focused on Landsat-8, Sentinel-2 and Sentinel-3.” *European Journal of Remote Sensing,* 48(1), 17-32.

Marden, B., Miller, T., and Richards, D. (2015). “Factors Influencing Cyanobacteria Blooms in Farmington Bay, Great Salt Lake, Utah.” A Progress Report of Scientific Findings From the 2013 Growing Season, The Jordan River/Farmington Bay Water Quality Council.

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J., and Lim, T. K. (2006). "A Landsat surface reflectance dataset for North America, 1990-2000." *Geoscience and Remote Sensing Letters IEEE,* 3(1), 68-72.

Matthews, M. W. (2011). "A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters." *International Journal of Remote Sensing,* 32(21), 6855-6899.

McCullough, I. M., Loftin, C. S., and Sader, S. A. (2013). "Landsat imagery reveals declining clarity of Maine's lakes during 1995-2010." *Freshwater Science,* 32(3), 741-752.

Nelder, J. A., and Wedderburn, R. W. M. (1972). "Generalized linear models*." Journal of the Royal Statistical Society,* 135(3), 370-384.

Olmanson, L. G., Bauer, M. E., and Brezonik, P. L. (2008). "A 20-year Landsat water clarity census of Minnesota's 10,000 lakes." *Remote Sensing of Environment*, 112(11), 4086-4097.

Page, B. P., Kumar, A., and Mishra, D. R. (2018). "A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom." *International Journal of Applied Earth Observation and Geoinformation,* 66, 69-81.

Palmer, S. C., Kutser, T., and Hunter, P. D. (2015). "Remote sensing of inland waters: Challenges, progress and future directions." *Remote Sensing of Environment,* 157, 1-8.

Pekel, J. F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). "High-resolution mapping of global surface water and its long-term changes." *Nature,* 540(7633), 418.

Recknagel, F., Orr, P., Swanepoel, A., Joehnk, K., and Anstee, J. (2018). "Operational forecasting in ecology by inferential models and remote sensing." *Ecological Informatics*, Springer, 319-339.

Rushforth, S. R., and Squires, L. E. (1985). "New records and comprehensive list of the algal taxa of Utah Lake, Utah, USA." *The Great Basin Naturalist,* 45(2), 237-254.

Rushforth, S. R., St. Clair, L. L., Grimes, J. A., and Whiting, M. C. 1981. "Phytoplankton of Utah Lake." *Great Basin Naturalist Memoirs,* 5(6), 85-100.

Søndergaard, M., Jensen, J. P., and Jeppesen, E. (2003). "Role of sediment and internal loading of phosphorus in shallow lakes." *Hydrobiologia*, 506(1-3), 135-145.

Stadelmann, T. H., Brezonik, P. L., and Kloiber, S. (2001). "Seasonal patterns of chlorophyll a and Secchi disk transparency in lakes of East-Central Minnesota: Implications for design of ground-and satellite-based monitoring programs." *Lake and Reservoir Management*, 17(4), 299-314.

Strong, A. E. (1974). "Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah Lake." *Remote Sensing of Environment,* 3(2), 99-107.

Tellman, B., Sullivan, J., Kettner, A., Brakenridge, G., Slayback, D., Kuhn, C., and Doyle, C. (2016). "Developing a global database of historic flood events to support machine learning flood prediction in Google Earth Engine." *AGU Fall Meeting Abstracts*.

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., and Nõges, T. (2016). “First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery.” *Remote Sensing*, 8(8), 640.

Torbick, N., Hession, S., Hagen, S., Wiangwang, N., Becker, B., and Qi, J. (2013). “Mapping inland lake water quality across the Lower Peninsula of Michigan using Landsat TM imagery.” *International Journal of Remote Sensing*, 34(21), 7607-7624.

[USEPA] US Environmental Protection Agency. (2017). “Recommendations for cyanobacteria and cyanotoxin monitoring in recreational waters.” *EPA/820/R-17/001*.

[USGS] US Geologic Survey: Utah Water Science Center. (2013). "Great Salt Lake—salinity and water quality." <https://ut.water.usgs.gov/greatsaltlake/salinity/> (April 5 2017).

[UDEQ] Utah Department of Environmental Quality. (2006). “Utah Lake.” *Watershed Management Program Lake Reports.*

Whiting, M. C., Brotherson, J. D., and Rushforth, S. R. (1978). "Environmental interaction in summer algal communities of Utah Lake." *Western North American Naturalist*, 38(1), 31-41.

Wurtsbaugh, W. A. (2008). "Nutrient loading and eutrophication in the Great Salt Lake." *Watershed Sciences Faculty Publications*, Paper 299.

Wurtsbaugh, W. A., Marcarelli, A., and Boyer, G. (2012). “Eutrophication and metal concentrations in three bays of the Great Salt Lake (USA).” *Final Report to the Utah Division of Water Quality*, Salt Lake City, Utah.

CHAPTER 4

A DATA-DRIVEN FORECASTING FRAMEWORK TO SUPPORT PROACTIVE MONITORING STRATEGIES

4.1 Introduction

Harmful algal blooms (HABs) have received an increasing amount of attention in both research and management communities following several years of widely-publicized bloom events that have occurred throughout inland waters of the United States (Michalak et al. 2013, Brooks et al. 2016, Brooks et al. 2017). Inland blooms can be harmful by producing toxins that pose an ecological and public health risk, by blocking light for other aquatic plants, and by creating hypoxic conditions when they decompose. When these blooms occur, water management entities are forced to address the economic and social impacts that blooms have on drinking water sources, treatment strategies, public health, recreation, and the environment. Over the past two decades, the paradigm of surface water quality monitoring and management, particularly concerning HABs, has been shifting from reactive to proactive (Schofield et al. 1999, Anderson et al. 2012). As this shift occurs, monitoring and regulatory agencies look to models and forecasts to anticipate future conditions so they can allocate resources and plan for monitoring activities, communicate with the public in a timely manner, and mitigate blooms (Twigt et al. 2011, Qin et al. 2015, Stumpf et al. 2016, Recknagel and Michener 2017).

Physically-based models representing hydrodynamic processes and weather conditions have been used to produce near-term forecasts of bloom conditions (such as the National Oceanic and Atmospheric Administration (NOAA) HAB Bulletins for Lake Erie). These types of models can be computationally expensive and require extensive information about lake characteristics (e.g. bathymetry, hydrodynamic patterns) and internal processes (e.g. species, nutrient loading, growth, and grazing rates). They can also be limited in their forecasting scope, generally providing projections on a daily or weekly basis. An alternative approach to forecasting HABs is a data-driven, statistical approach, such as with a Bayesian Networks (BN). A BN approach has a number of advantages: it allows for integration of many processes over varying spatial or temporal scales (not necessarily limited to a single computational timestep), and while other statistical models (e.g. multivariate linear regression) cannot handle missing observations, BNs can use incomplete datasets. In many regions, historical data that are needed for model calibration (measures of blooms and predictive variables) are often very limited or are incomplete. BNs are also effective tools for communication with nontechnical users through a graphical representation of inputs and relationships and uncertainty (Jensen 2001, Obenour et al. 2014, Scutari and Denis 2014).

Applications of algal bloom-related BNs include estuarine models of chlorophyll *a* (chl- *a*) (Alameddine et al. 2011) and other algae measures (Nojavan et al. 2014), HAB-related water quality indices in rivers (Forio et al. 2015), and phytoplankton biomass in lakes (Malve et al. 2007). These examples generally approach the challenge of predicting blooms as a function of other water quality constituents and characteristics (e.g. nutrient concentrations, temperature) and physical processes (e.g. radiance, growth rates) which could also be represented in a process-based, hydrodynamic model. Some BNs, however, have been used to explore water quality from a more external perspective that has implications for watershed and water resource management activities. An example of this is the evaluation of nutrient management strategies in a Utah reservoir, which represents the physical watershed and reservoir (including economic and social aspects) as a BN in order to evaluate management alternatives (Ames et al. 2005).

This study presents a hydroinformatic framework built around a BN for forecasting measures of algal bloom-related water quality conditions on a seasonal basis. The framework couples readily-available hydrology and climate data with the predictive BN model to forecast probable water quality conditions and inform water quality monitoring and management decisions. This study 1) provides an overview of the development of a seasonal BN model, 2) outlines a generic framework using open-source software and web services, and 3) demonstrates (through hindcasting) the framework to show how a forecasting framework can support water quality monitoring and management decisions. This framework is demonstrated for Utah Lake, which has a long history of algal blooms; however, it is broadly applicable to other lakes.

4.1.1 Study Area

4.1.1.1 Overview

Utah Lake is a large, shallow lake with a surface area of approximately 380 km^2, an average depth 3.2 m, and maximum depth of 4.3 m. The designated beneficial uses of Utah Lake include: infrequent primary contact recreation, warm-water game fish, other aquatic wildlife, and agricultural uses (irrigation and stock watering). The lake receives treated wastewater and stormwater from rapidly growing urban areas near the lake, irrigation runoff, and a number of springs. It is directly linked to the Great Salt Lake as water drains from Utah Lake via the Jordan River, which is also a receiving water for treated wastewater and other urban outputs, and streamflow from a series of canyons to the east of the lake system. The entire lake system provides habitat for millions of migratory birds (Cox and Kadlec 1995) and has major recreational value with several state parks, marinas, and campgrounds along the shores.

In 2016, a major bloom occurred in Utah Lake that prompted major public concern and was severe enough that irrigation water providers were forced to switch to alternative sources. The ripple effect of the bloom extended to municipal water providers who had to adjust operations to meet irrigation needs and use more expensive sources for drinking water supplies. During this event, cell density for cyanobacteria reached levels greater than 100,000 cells/mL and microcystin (a cyanotoxin that can affect the liver and other internal organs, causing nausea, diarrhea, and vomiting) were found at several sites throughout the lake (Utah DEQ 2016). Water quality concerns for this lake system are not new; in 2004, the lake was listed as impaired due to exceedances of state criteria for total phosphorus (TP) and total dissolved solids (TDS) concentrations (SWCA 2007). Historical records of algal biomass (estimated via field-measured chlorophyll *a*) are as high as 325 µg/L. In the past, efforts to study and forecast blooms have typically focused on lakes that provide drinking water (such as Deer Creek Reservoir in Utah or Lake Eerie in Ohio). However, many of the negative impacts of blooms including degraded aesthetics and odors, health risks from recreational contact, and decreasing dissolved oxygen are also relevant to nondrinking water sources (such as Utah Lake). These impacts have been widely explored and documented in a variety of regions (Paerl et al. 2001, Anderson 2002, Landsberg 2002, Heisler et al. 2008).

4.1.1.2 Influences to Utah Lake Hydrology and Water Quality

An important defining feature of Utah Lake is that it is a receiving body for several snowpack-dominated streams and watersheds in the growing urban area of Provo-Orem (located about 40 miles or 70 km south of Salt Lake City). The largest inflows to the lake come from two major streams, the Provo River and Spanish Fork River (together, these account for approximately 60% of the total inflow).

Utah Lake acts as a phosphorus sink, with estimates of annual phosphorus loading ranging between 270 and 300 tons from wastewater, streams, groundwater, and springs (Merritt 2017, SWCA 2007). Estimates of additional phosphorus from atmospheric deposition vary widely from between 18-380 tons/year (Olsen 2018) to 1600 tons/year (Merritt 2017). Because of the abundance of phosphorus in the lake and its sediments, simple relationships between external nutrient loading and algal blooms are difficult to quantify. This has led some to suggest that blooms in Utah Lake are largely triggered or limited by mixing, turbidity (Merritt 2017), and resuspension of sediment-bound phosphorus (Abu-Hmeidan et al. 2018). Hydrologic inputs to the lake (including precipitation and streamflow from the Provo and Spanish Fork Rivers) can contribute to some of these hydrodynamic processes.

4.2 Methodology

4.2.1 BN Model Development

The workflow used for developing the BN model for Utah Lake involved data collection and discretization, variable and structure specification, calculation of conditional probabilities, and model validation.

<u>4.2.1.1 Data Collection and Discretization</u>

4.2.1.1.1 Defining an Indicator of Potential Harmful Algal Blooms

Algal bloom-related water quality conditions were derived from the historical remotely-sensed record of chl- *a* for Utah Lake. This record contains estimates of chl- *a* for the entire lake (at a 30-meter resolution) from 1984-2016. While chl- *a* concentrations do not indicate whether a bloom is made of cyanobacteria or if the cyanobacteria are actively producing toxins, it remains one of the most consistent measures of algal blooms and has been used for decades to evaluate trophic state of lakes (Thiemann and Kaufmann 2000, Duan et al. 2007) via the Carlson Trophic State Index (TSI) (Carlson 1977). Potential impacts to lakes in the context of chl- *a* and trophic states are shown in Table 4.1.

The maximum of chl- *a* concentrations for the entire lake were calculated for each image in the historical record, and then summarized by month. This monthly maximum serves as a measure of extreme algae biomass that can be modeled by the BN. Observations were then discretized, or transformed, into categorical data. To provide an example of a possible threshold of interest to monitoring agencies, the monthly lake maxima data were categorized using a threshold of 275 µg/L. The 275 µg/L threshold corresponds with the top 25th percentile of chl- *a* concentrations. At higher chl- *a* concentrations, there is generally a greater risk for cyanobacteria blooms and algal scums that would impact recreation, aesthetics, and aquatic organisms as they decompose. Depending on the needs of the monitoring agency and characteristics of the lake, this threshold can be used as a useful indicator of critical conditions, which can guide monitoring decisions (e.g. trigger sampling). Because of the heightened risk of cyanobacteria, this threshold can trigger additional monitoring for cyanobacteria cell density or potential cyanotoxins. These other measures are rarely collected unless there is a suspected or observed bloom. This level can also indicate the need to issue an advisory for recreation. The distributions of monthly average and maximum chl- *a* estimates are shown in Figure 4.1 to illustrate the range of past conditions.

4.2.1.1.2 Predictive Variables

Selection of predictive factors was limited based on a number of constraints: they must be representative of influences observed in literature (i.e. inflows and contributors to lake mixing (Page et al. 2018)), records must be concurrent with the remotely sensed record (1984-2016), and data for these variables must be currently available through web services for monitoring or forecasting programs including the United States Geological Survey National Water Information System (USGS NWIS), the Natural Resources Conservation Service Snow Telemetry (NRCS SNOTEL) network, and the National Oceanic and Atmospheric Administration National Climate Data Center (NOAA NCDC). For Utah Lake, this limits variables to daily streamflow (cfs) from the USGS NWIS database for the two largest stream inflows to the lake: Provo River and Spanish Fork River (Gage 10163000 and 10150500, respectively), daily precipitation (mm/day) and air temperature (degrees F) from NOAA NCDC for weather station USC00427064, and daily snow water equivalent (SWE) (in) from the SNOTEL station 820: Timpanogos Divide.

Previous studies of Utah Lake and other lakes have suggested winter and spring hydrology and climate as major contributing factors for algal blooms (Michalak et al.

2013, Page et al. 2018). However, the scale and lagged effect of these variables are not certain. Scale refers to the level of aggregation (e.g. spring versus monthly totals), while lagged effect refers to the latency between when a variable is measured or observed and when its effects are manifest in a bloom event. For example, based on the study of a single bloom that was concentrated in July (Page et al. 2018), it is not clear whether the winter climate conditions are as useful predictive factors for water quality conditions in September as they are for predicting conditions in July. To explore the potential of a variety of scales, the daily time series for each of the variables were aggregated to both monthly and seasonal totals or averages. Seasons are defined as winter (October-March) for SWE, and spring (March-May) and summer (June-August) for all other variables.

There are several approaches that are commonly used for discretization of continuous data, and it has been shown that the chosen method can have an impact on the results of the BN (Nojavan et al. 2017). Discretization involves defining both a number of categories or states that the continuous data can be classified into and defining thresholds for these categories. Generally, a categorical variable with fewer states will result in a BN that is simpler and easier to understand. Thus, for the relationships between predictive factors and measures of water quality to be clear and easy to understand, we limited discretization to two categories (above and below a single threshold). The "optimal" threshold for a given variable was determined by comparing predictive ability for different thresholds ranging from 25^{th}-75^{th} percentile. First, a variable was categorized as either "high" or "low" if it were above or below a given percentile. Then, the mutual information (MI) criteria between the discretized variable and the TSI in July-September was calculated. The MI describes how much variance of a given variable is described by its parent or child, or how much knowing the state of one variable can reduce the uncertainty of the other (if the two variables have no relation at all, the MI is zero). The MI for two variables, X and Y, can be calculated as follows:

$$I(X,Y) = \sum_{y \in Y} \sum_{x \in X} P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right) \qquad (4.1)$$

where P(x,y) is the joint probability of X and Y with respective states x and y, and P(x) and P(y) are the marginal probabilities of X and Y. This process was repeated for each variable and both seasonal and monthly scales. The mutual information (MI) criteria was also used to compare relative importance of different variables (similar to Ames et al. 2005). Variables and scales with the highest relative MI were included as predictive variables in the model.

4.2.1.2 Defining BN Model Structure

The structure of the BN consists of nodes and arcs, with nodes representing each variable, and the arrows connecting the nodes (referred to as arcs) indicate the direction of dependence between the nodes. The nodes and arcs are represented in a Directed Acyclic Graph (DAG). The term DAG describes three important characteristics of the model: 1) the relationships are directed, or there is some sense of causality or dependence from the parent nodes to the children nodes; 2) arc cannot result in a cycle or feedback loop within the model; and 3) the relationships can be represented visually, making the BN an effective tool for communicating the relationships contained within the model. For this study, the structure of the DAG (nodes and arcs) was specified using R Statistical Software and the *bnlearn* package (Scutari, 2010).

4.2.1.3 Calculating Conditional Probabilities

The joint probability distribution representing the probabilities of all states of the variables in the network (also called the global probability distribution) is specified through a series of local distributions and conditional probability tables (CPT). The CPT reflects the probability of the states of a child node that is dependent on the state or condition of the parent node. In other words, the CPT defines the conditional dependence between a child and its parent node(s). These CPTs were populated using the *bnlearn* package and frequencies observed in the historical records.

In general, the probability of state x for a child node X with parent node Y having state y can be calculated as follows:

$$\mathrm{P}(X = x|Y = y) = \frac{P(X=x,Y=y)}{\mathrm{P}(Y=y)} \tag{4.2}$$

4.2.2 Forecasting Framework

A forecasting framework, which is built around the BN model, is suggested as a useful means of linking seasonal hydrologic and climate conditions with monitoring and management decisions (Figure 4.2). The generic framework involves establishing a predictive model (the BN) that relates the predictive variables to measures of water quality. The framework requires a connection to external databases and modeling applications. Web services are a convenient way to connect to and retrieve data from these sources in an automated fashion, with minimal parameters (usually elements such as parameter codes, dates of interest, and site identifiers). These data are retrieved and input into the predictive model as evidence. The web services for each database or tool return the data as a time series, which is then aggregated to the appropriate scale. The aggregated data are then classified according to the thresholds set in BN model development and used as evidence for the BN model. Those variables with available data are used as evidence to calculate the joint conditional probability of the worst state of a water quality condition, or likelihood that the worst state will occur. If no data are available for a particular variable by a certain date (for example, no data would be available for a June precipitation variable prior to the month of June), then this variable would not be used as evidence in calculating the joint probability. Finally, the likelihood of an event must be communicated to monitoring or management entities in order to inform monitoring activities or advisory decisions for recreation.

For Utah Lake, the framework was implemented within a simple web application built using the *Shiny* package in R. The application contains a server script that retrieves the most recent seasonal or monthly data (based on the current date and availability of data) from external databases using web services and existing R packages. The USGS NWIS database provides streamflow data for the Provo and Spanish Fork River, which are accessed using the *dataRetrieval* package (developed by the USGS). Precipitation and air temperature data are obtained from the NOAA NCDC database using the *rnoaa* package (developed by ROpenSci). The NRCS Air and Water Database (AWDB) houses SNOTEL data; the *RNRCS* package (developed by Robert Lee and Josh Roberti) is used to obtain SWE data.

The snowpack-driven hydrology of the study area enables early forecasting of lake chl- *a* conditions. Using SWE observations as evidence, the automated data retrieval enables preliminary forecasts for July and August conditions as early as April. By June, forecasts can be updated because spring streamflow is either known via streamflow gages or an informed estimate using air temperature and SWE observations. Forecasts continue to be refined as other variables become known and their data can be automatically retrieved and used as evidence in inference queries.

The most updated probability of high chl- *a* condition is relayed to the user via a graphical user interface, along with accompanying recommendations for monitoring and advisory actions.

4.2.3 Demonstrating the Framework through Retrospective Application

To demonstrate the utility of this framework and its potential impact on monitoring decisions, the model is applied over a historical period. The "hindcasted" recommendations for sampling are then compared to the actual sampling activity. For this retrospective application, historical observations of hydrologic and climate conditions are used as inputs into the BN model. Approximate inference using the *cpquery* function in the *bnlearn* R package is used to calculate the probability that a critical event (defined as chl- *a* concentrations above the threshold) would occur for each month. If in a given month the critical event was more likely than not to occur (i.e. has a probability greater than 50%), then there would be a recommendation would for frequent sampling during that month. The approximate inference is applied for the months of July, August, and September over the period of 1995-2014. This period corresponds with the existing UDWQ field sampling record in Utah Lake. Recommendations using the historical observations of external factors as evidence are then compared to actual monitoring practices (whether or not sampling was performed).

4.3 Results

4.3.1 BN Model

The comparison of MI criteria for predictive variables allowed comparison of relative importance of predictive variables for each month's conditions. A comparison is provided for the "optimal" thresholds for spring streamflow, monthly precipitation, and previous month's water quality conditions in Table 4.2. In July and August, spring streamflow was the most important predictor of chl- *a* state, with the highest MI. However, spring streamflow had a relatively low MI for September chl- *a* and previous conditions from August had greater relative importance. The CPTs help describe the extent and lag of influence of variables; streamflow is less influential to water quality conditions late in the summer and precipitation for the preceding month influential, but earlier precipitation was not as useful in predicting chl- *a* state.

Though SWE was not as good of a predictor of chl- *a* as spring streamflow, the two variables are highly correlated. SWE was included in the model as a predictive variable for spring streamflow, which enables earlier predictions (total SWE is available by March, while observations for spring streamflow will not be available until June). Similarly, air temperature was a consistently poorer predictor for monthly chl- *a* (lower MI) at both monthly and seasonal scales. However, it was a significant predictor for spring streamflow.

The final DAG is shown in Figure 4.3. The DAG visually conveys the lagged nature of the effects, with winter SWE and spring temperatures influencing spring streamflow, and the spring streamflow, earlier summer bloom conditions, and precipitation totals in preceding months impact the bloom measures in later months.

Examples of the CPTs for the monthly chl- *a* conditions and streamflow are shown in Figure 4.4. Some of the relationships observed in the CPTs follow what was observed in the 2016 July bloom; that is, when there is high spring streamflow and low precipitation, there is a greater likelihood of higher chl- *a* concentrations in the early summer. SWE and streamflow are highly correlated, with below average SWE likely to produce below average spring streamflow.

An independent validation of model performance is not practical due to the limited number of observations. Model performance was therefore evaluated by using the historical observations of predictor variables as evidence and historical observations of chl- *a* as events. The *cpquery* function in the *bnlearn* package estimates the conditional probability of an event occurring (or measure of water quality having a certain state) given some evidence. For example, in 1986, there was above average spring streamflow, and low June precipitation. For this case, the conditional probability of a critical event (chl- *a* above the threshold) would be calculated using evidence that the spring streamflow was above average and June had low precipitation. In some years, not all parent nodes had observations, so only factors with available data were used as evidence (e.g. in 2016, August precipitation data were not available, so the only evidence used to predict the states of chl- *a* condition for September were the states of August chl- *a*). This process is repeated for all years in the historic record, and the number of years where the event was predicted to be more likely than not to occur (>50% probability) were counted as "correct" predictions. The BN model correctly predicted occurrence of events for maximum chl- *a* states between 81-91% of years (n=32) depending on the month. There are approximately equal numbers of false positives as false negatives over all of the observations. This indicates that there is no clear bias towards overestimation (predicting high likelihood for a critical event when none was observed) or underestimation (predicting low likelihood for a critical event when one was observed). Exploration of the factors that were common among the incorrect predictions identifies which predictive variables may have greater uncertainty or may have the greatest tendency to incorrectly predict chl- *a* state. Above average spring streamflow was common among overestimations, while below average streamflow was generally common among underestimations. This may indicate a higher level of uncertainty or weakness in the predictive power of spring streamflow, despite its relatively high MI.

4.3.2 Demonstrating the Framework: Retrospective Monitoring Decisions

The retrospective sampling recommendations from 1995-2014 compared to the actual historical sampling practices are shown by month in Figure 4.5. These results indicate that had this framework been in use (with a minimum threshold of 50% likelihood of chl- *a* greater than 275 µg/L), there would have been a slight increase of monitoring activity. Over the 60 months in the hindcasting period (from July-September of 1995-2014), there were 37 months with samples. The retrospective application, however, recommended sampling during 43 of these months. There were 18 months where there was a recommendation according to the hindcast, but no sample was collected. The hindcast also shows that there were 21 months where samples were collected but there was no recommendation for sampling. As the threshold of acceptable risk is increased, the recommended sampling protocol is further refined. For example, with a threshold of 60% likelihood of high chl- *a*, there are a total of 39 months with sampling recommendations, 12 months where there was a recommendation but no sample collected, and 12 months with no recommendations but samples were collected. The results indicate that frequency of monitoring may not need to drastically increase, but a simple forecasting tool could help focus sampling activities during months where risks for poor conditions are particularly high and more frequent monitoring is especially critical.

4.4 Discussion

The DAG for Utah Lake reflects observed patterns where seasonal hydrological and climate variables have relationships to bloom conditions throughout the summer, but also that the relationships vary in scale and latency depending on the variable. The precipitation total from the preceding month appeared to have a relatively immediate and straightforward effect, while the streamflow was a bit more lagged, with conditions in July and August responding to streamflow from earlier in the spring rather than streamflow from the month immediately preceding. These patterns follow what was found by the Page et al. study of the 2016 bloom – there are observed relationships between inflows and precipitation early in the year and algal biomass later in the year. Additionally, seasonal and monthly higher temperatures were found to be a relatively poor predictor for chl- *a* concentrations during July-September; however, analysis of the immediate effects of high temperatures indicates a general increase in chl- *a* concentrations. This suggests that the effects of external factors, the scales at which they are measured, and the latency of their influence must be considered when forecasting.

The relative importance of streamflow on Utah Lake has implications for water resources beyond immediate monitoring and advisory actions on the lake. The higher streamflow tied to greater probabilities of high chl- *a* concentrations suggests that as the surrounding watersheds continue to develop and become more urbanized, there is likely to be an increase in stormwater runoff, and a subsequent increase in streamflow. Engineered projects that affect inflows to the lake, such as delta and wetland restoration, should also be taken into account. These types of alterations could impact the degree to which streamflow influences water quality or the lag between measures of streamflow and when high chl- *a* concentrations occur, requiring adaptation of the model structure or conditional probabilities tables.

4.4.1 Potential and Challenges for Scaling

The presented framework for forecasting is suggested as a general approach that may be scaled to other regions. The framework must be general to allow for the potential variations in lake characteristics, inputs, and measures of interest.

The processes and factors that contribute to lake algal blooms are complex and may differ from one lake to another. The BN structure in this study is able to represent influences of external factors on varying scales (some seasonal, some monthly), and these effects may present themselves differently in other areas. For example, water quality conditions in a lake that does not receive snowmelt-driven streamflow may be less sensitive to spring streamflow than those in Utah Lake. However, the general processes (inflows from contributing watersheds and occurrence of large storms) follow what has been documented in literature for this lake and others (Michalak et al. 2013, Page et al.

2018). The general principle of the framework, using external hydrological and climate data to forecast water quality conditions, may be generally applied if the scales and variables representing hydrologic and climate variables are adapted to reflect the unique characteristics of other different regions.

Another principle of the framework that suggests potential for broader scaling is the use of web services that connect to national databases, meaning that this framework could apply to any lake with nearby streamflow and weather gages. For Utah Lake, this included data from NOAA and USGS data; however, there are many state or regional databases that could provide additional measurements, provided there are means of automatically retrieving data. For areas without nearby gages or weather stations to provide observational data, there are other methods for obtaining evidence. This can be done through forecasting models, such as the National Water Model, which can provide streamflow estimates with national coverage, or NRCS SNOTEL snowpack and SWE forecasts for large areas of the United States.

Some challenges that may inhibit scaling of the framework to a regional or national level involve the limited nature of historical bloom or water quality measures. In many lakes and reservoirs, this information may not be readily available. Remote sensing, as was used to provide the records used in this study, is an option for constructing a more regular and robust record, provided that there are reliable models available or sufficient data for calibrating models.

When scaling this framework to other regions, there will likely be differences in measures of algal bloom-related water quality. A lake with generally lower trophic states may require a different chl- *a* threshold than that used for Utah Lake, which generally fluctuates between eutrophic and hypereutrophic so that distinctions are between "poor" and "very poor" conditions. A specific example of this might be for Florida waters, where the specific chl- *a* threshold of 40 µg/L is widely used to designate bloom conditions (though even in areas where there are established thresholds, alternative measures may still be used (Bachmann et al. 2003)). Other measures of algal blooms may also be more appropriate for other areas; for example, if cyanobacteria cell density data, other pigments, or toxin concentrations are available for other regions, these could be used in place of chl- *a* and TSI to target HABs. Additionally, not all lakes may be as sensitive to these particular hydrologic and climate loads (instead, they may be much more affected by stratification and lake turnover, waste water inputs, or other sources of nutrients).

4.4.2 Limitations and Additional Considerations

It is important to acknowledge that there are other parameters that are of particular interest to this lake system but were not included in this study including: agricultural runoff from the surrounding watersheds, wastewater effluent from municipalities, and nutrient loading from these sources. These variables are especially important because of the ties to watershed and waste water management actions; however, these data are often not available (e.g. watershed runoff would require a calibrated model), are limited in the extent of the historical record of chl- *a*, or are not available for automated data retrieval and coupling to a predictive model.

Additionally, other methods for structuring the model could also be considered. For example, some programs (including *bnlearn*) allow the use of continuous data and increasing the possible number of states may provide additional information about influence of a given variable. These alternatives do have drawbacks, however, such as increasing the complexity of the conditional probabilities. There are also alternative methods for representing the variables such as binary indicator variables of an event (e.g. whether there was an intense storm event) or as departure from the mean condition may provide additional insight into how these external factors contribute to blooms.

Other important considerations for this framework include the uncertainty of the model variables, the relationships, and the forecast results. The BN model conveys uncertainty in the variables through their marginal distributions, the relationships through the conditional probabilities, and the results by providing the likelihood of a condition occurring rather than a deterministic value. The marginal distributions and conditional probabilities describing each of the variables and their relationships may need to be updated in the case shifts due to changing climate or hydrologic patterns (e.g. if new rainfall patterns occur, the marginal distribution of monthly precipitation nodes will need to be updated). Additionally, the level of acceptable risk or uncertainty in the result may vary depending on factors such as available resources (as resources or personnel are more limited, the threshold for risk may increase so that resources are used only when there is a higher risk for poor conditions) or levels of public exposure (if there is greater volume of recreation during certain months, this may lower the threshold of acceptable risk).

4.5 Conclusion

The framework for modeling and forecasting presented in this study has the potential to benefit monitoring and lake management entities and help them shift towards more proactive strategies. This approach to monitoring addresses a common issue in water quality monitoring where "you only find what you are looking for," by helping monitoring entities anticipate when they should be "looking" for something. There is evidence from the remotely sensed records that poor conditions occurred in the past when no samples were collected, so the conditions were not recorded by any monitoring agency.

The hindcasting demonstration illustrated how historical monitoring frequency may have been insufficient to capture many poor water quality conditions, but in the future, this framework could be used to plan ahead for increased monitoring. In some cases, the model could also conserve resources, if sampling is only performed when there is a high risk for poor conditions. The forecasts provided by this framework can be used to guide monitoring strategies, with more focused allocation of personnel and resources for sampling and monitoring when severe blooms are likely. This framework has the potential to be scaled by exploiting readily available data sources and web services, reflecting local hydrology and climate influences, adopting alternative algal bloom measures, and using appropriate levels of risk so that it can support and improve water quality monitoring and management practices in other regions.

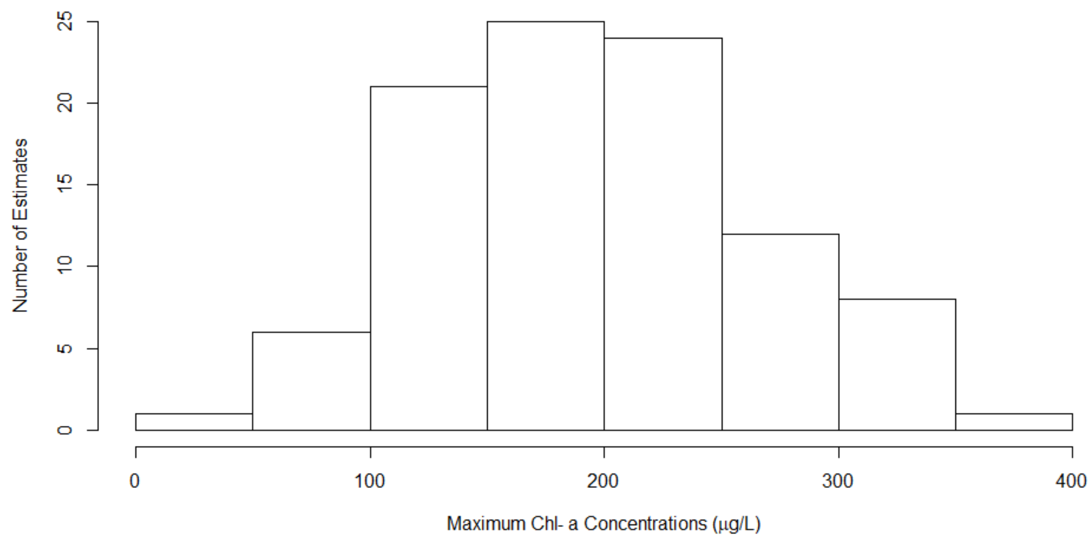

Figure 4.1 Distribution of Monthly Maximum Lake Chl- *a*

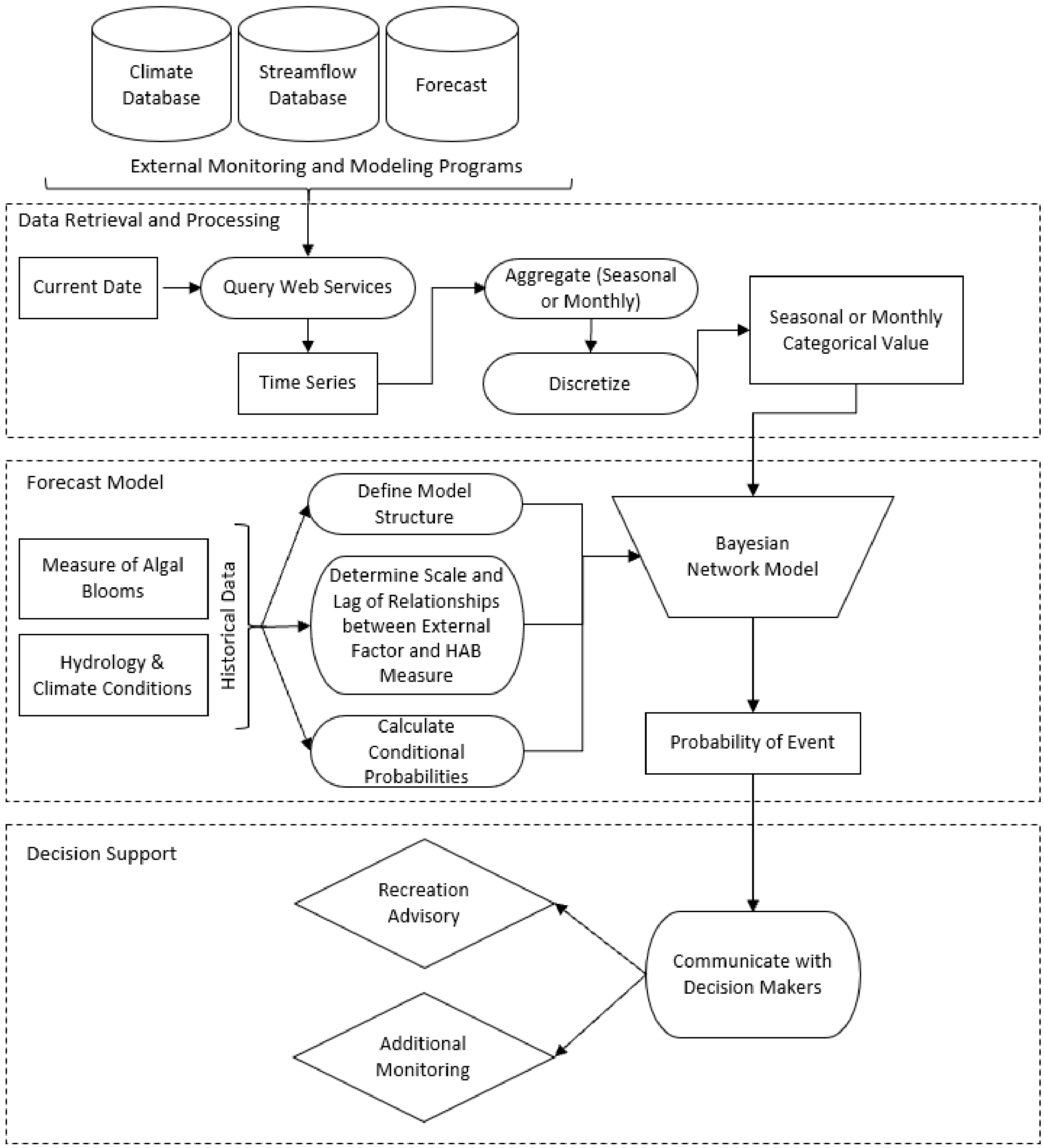

Figure 4.2 Decision Support Framework Using the BN and Automated Data Retrieval from External Databases

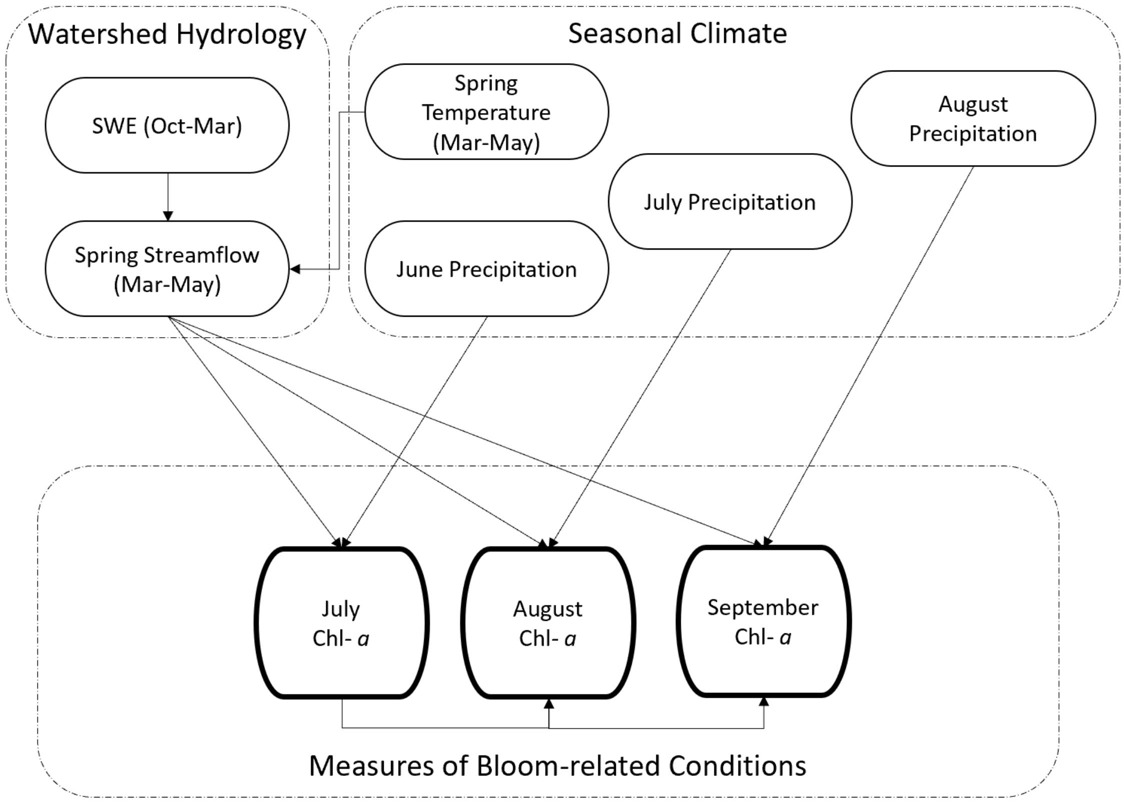

Figure 4.3 Directed Acyclic Graph of the Utah Lake BN Model

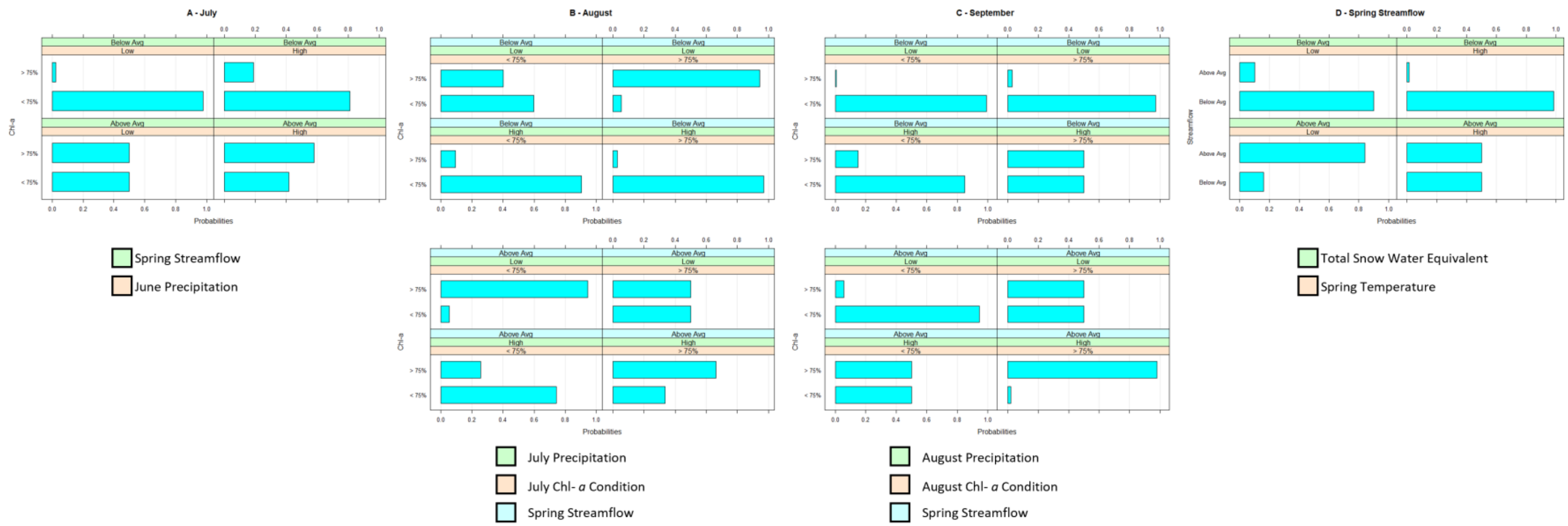

Figure 4.4 Conditional Probability Tables for Monthly Extreme Chl- *a* Measures from A) July, B) August, and C) September, and D) Streamflow

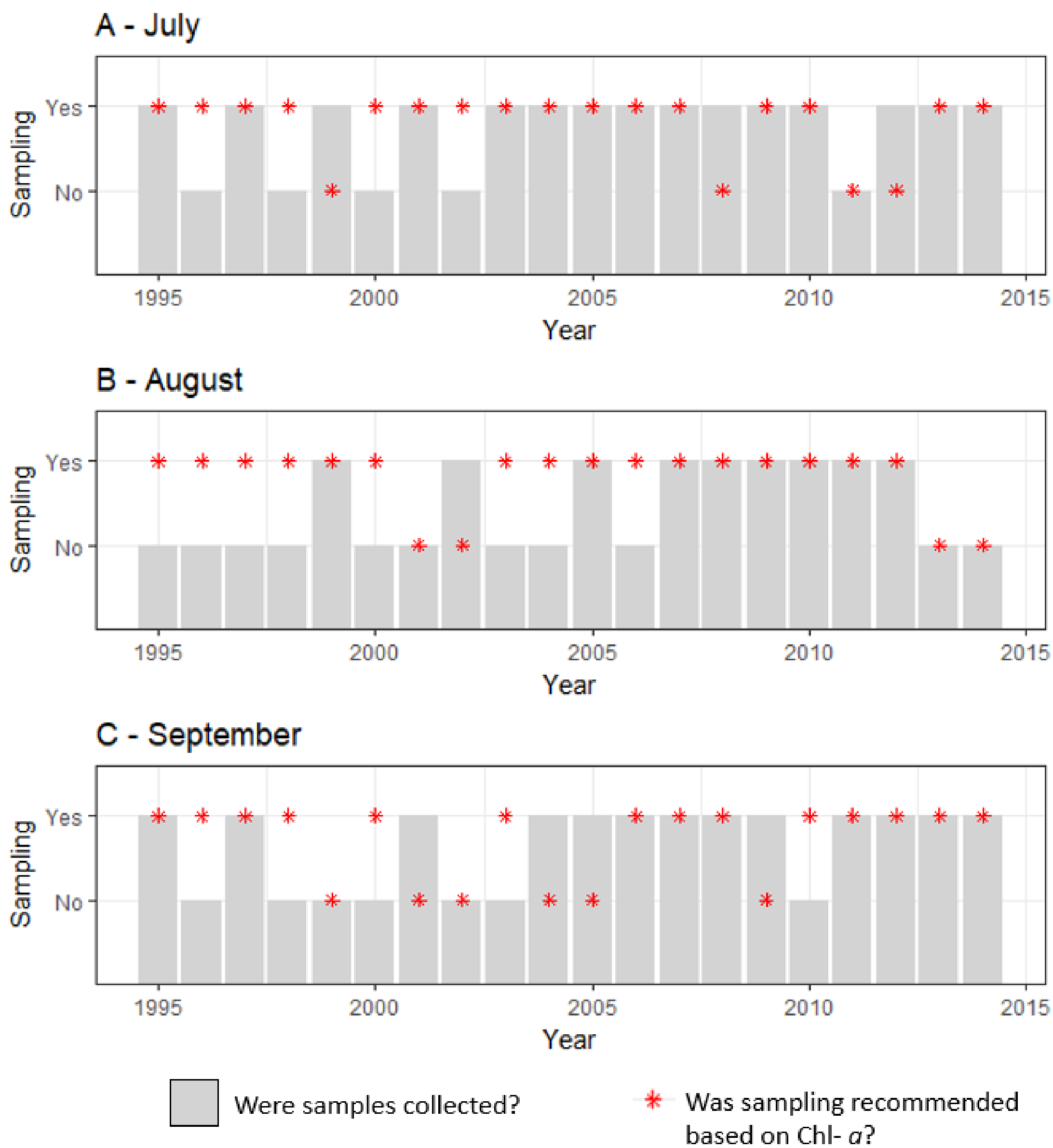

Figure 4.5 Comparison of Historical Sample Collection Versus the Recommended Sampling Based on the Historical Application of the BN Model for A) July, B) August, and C) September.

Table 4.1 Water Quality Impacts Related to Trophic State and Chl- *a*

Chl- *a* Concentration (μg/L)	**Carlson Trophic State Index**	**Trophic Classification**	**Potential Impacts to Lake**
<7	<50	Oligotrophic - Mesotrophic	
7-20	50-60	Eutrophic	Hypolimnion may be at risk for anoxic conditions
20-56	60-70		Blue-green algae are likely to dominate, algal scums are possible
56-155	70-80	Hypereutrophic	Dense algae
>155	>80		Severe algal scums, Fish kills possible

Table 4.2 Summary of MI for Variables Used in Final Model

Monthly TSI	Spring Streamflow	Previous Month's Precipitation	Previous Month's TSI
July	4.9	3.5	-
August	6.6	2.7	1.5
September	2.0	4.0	3.7

4.6 References

Abu-Hmeidan, H. Y., Williams, G. P., and Miller, A. W. (2018). "Characterizing total phosphorus in current and geologic Utah Lake sediments: implications for water quality management issues." *Hydrology*, 5(1), 8.

Alameddine, I., Cha, Y., and Reckhow, K. H. (2011). "An evaluation of automated structure learning with Bayesian networks: an application to estuarine chlorophyll dynamics." *Environmental Modelling & Software*, 26(2), 163-172.

Ames, D. P., Neilson, B. T., Stevens, D. K., and Lall, U. (2005). "Using bayesian networks to model watershed management decisions: an East Canyon Creek case study." *Journal of Hydroinformatics*, 7(4), 267-282.

Anderson, D. M., Glibert, P. M., and Burkholder, J. M. (2002). "Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences." *Estuaries*, 25(4), 704-726.

Anderson, D. M., Cembella, A. D., and Hallegraeff, G. M. (2012). "Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management." *Annual Review of Marine Science*, 4, 143-176.

Bachmann, R. W., Hoyer, M. V., & Canfield Jr, D. E. (2003). “Predicting the frequencies of high chlorophyll levels in Florida lakes from average chlorophyll or nutrient data.” *Lake and Reservoir Management*, 19(3), 229-241.

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V. V., Morton, S. L., Perkins, D. A., Reavie, E. D., Scott, G. I., Smith, S. A., and Steevens, J. A. (2016). "Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems?" *Environmental Toxicology and Chemistry*, 35(1), 6-13.

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V. V., Morton, S. L., Perkins, D. A., Reavie, E. D., Scott, G. I., Smith, S. A., and Steevens, J. A. (2017). "In some places, in some cases, and at some times, harmful algal blooms are the greatest threat to inland water quality." *Environmental Toxicology and Chemistry*, 36(5), 1125-1127.

Carlson, R. E. (1977). "A trophic state index for lakes." *Limnology and Oceanography*, 22(2), 361-369.

Cox, R. R., and Kadlec, J. A. (1995). "Dynamics of potential waterfowl foods in Great Salt Lake marshes during summer." *Wetlands*, 15(1), 1-8.

Duan, H., Zhang, Y., Zhang, B., Song, K., and Wang, Z. (2007). "Assessment of chlorophyll-a concentration and trophic state for Lake Chagan using Landsat TM and field spectral data." *Environmental Monitoring and Assessment*, 129(1), 295-308.

Forio, M. A. E., Landuyt, D., Bennetsen, E., Lock, K., Nguyen, T. H. T., Ambarita, M. N. D., Musonge, P. L. S., Boets, P., Everaert, G., and Dominguez-Granda, L. (2015). "Bayesian belief network models to analyse and predict ecological water quality in rivers." *Ecological Modelling*, 312, 222-238.

Heisler, J., Glibert, P. M., Burkholder, J. M., Anderson, D. M., Cochlan, W., Dennison, W. C., Dortch, Q., Gobler, C. J., Heil, C. A., and Humphries, E. (2008). "Eutrophication and harmful algal blooms: a scientific consensus." *Harmful Algae*, 8(1), 3-13.

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*, Springer, New York, N. Y.

Landsberg, J. H. (2002). "The effects of harmful algal blooms on aquatic organisms." *Reviews in Fisheries Science*, 10(2), 113-390.

Malve, O., Laine, M., Haario, H., Kirkkala, T., and Sarvala, J. (2007). "Bayesian modelling of algal mass occurrences—using adaptive MCMC methods with a lake water quality model." *Environmental Modelling & Software*, 22(7), 966-977.

Merritt, L. B. (2017). "Utah Lake: A Few Considerations." <https://le.utah.gov/interim/2017/pdf/00004935.pdf> (July 1, 2018)

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., Chaffin, J. D., Cho, K., Confesor, R., and Daloğlu, I. (2013). "Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions." *Proceedings of the National Academy of Sciences*, 110(16), 6448-6452.

Nojavan, F., Qian, S. S., Paerl, H. W., Reckhow, K. H., and Albright, E. A. (2014). "A study of anthropogenic and climatic disturbance of the New River Estuary using a Bayesian belief network." *Marine Pollution Bulletin*, 83(1), 107-115.

Nojavan, F., Qian, S. S., and Stow, C. A. (2017). "Comparative analysis of discretization methods in Bayesian networks." *Environmental Modelling & Software*, 87, 64-71.

Obenour, D. R., Gronewold, A. D., Stow, C. A., and Scavia, D. (2014). "Using a Bayesian hierarchical model to improve Lake Erie cyanobacteria bloom forecasts." *Water Resources Research*, 50(10), 7847-7860.

Olsen, J. (2018). "Measuring and Calculating Current Atmospheric Phosphorous and Nitrogen Loadings on Utah Lake Using Field Samples, Laboratory Methods, and Statistical Analysis." *2018 J. Paul Riley AWRA Utah Section Student Paper Competition*. BYU Salt Lake Center.

Paerl, H. W., Fulton, R. S., Moisander, P. H., and Dyble, J. (2001). "Harmful freshwater algal blooms, with an emphasis on cyanobacteria." *The Scientific World Journal*, 1, 76-113.

Page, B. P., Kumar, A., and Mishra, D. R. (2018). "A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom." *International Journal of Applied Earth Observation and Geoinformation*, 66, 69-81.

Qin, B., Li, W., Zhu, G., Zhang, Y., Wu, T., and Gao, G. (2015). "Cyanobacterial bloom management through integrated monitoring and forecasting in large shallow eutrophic Lake Taihu (China)." *Journal of Hazardous Materials*, 287, 356-363.

Recknagel, F., and Michener, W. K. (2017). *Ecological Informatics: Data Management and Knowledge Discovery*, 3rd ed., Springer, Heidelberg, Germany.

Schofield, O., Grzymski, J., Bissett, W. P., Kirkpatrick, G. J., Millie, D. F., Moline, M., and Roesler, C. S. (1999). "Optical monitoring and forecasting systems for harmful algal blooms: possibility or pipe dream?" *Journal of Phycology*, 35(6), 1477-1496.

Scutari, M., and Denis, J.-B. (2014). *Bayesian Networks: With Examples in R*, CRC press, Boca Raton, FL.

Scutari, M. (2010). "Learning Bayesian Networks with the bnlearn R Package." *Journal of Statistical Software*, 35(3), 1-22.

Stumpf, R. P., Johnson, L. T., Wynne, T. T., and Baker, D. B. (2016). "Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie." *Journal of Great Lakes Research*, 42(6), 1174-1183.

SWCA. (2007). "Utah Lake TMDL: pollutant loading assessment & designated beneficial use impairment assessment," *Prepared for State of Utah Division of Water Quality*.

Thiemann, S., and Kaufmann, H. (2000). "Determination of chlorophyll content and trophic state of lakes using field spectrometer and IRS-1C satellite data in the Mecklenburg Lake District, Germany." *Remote Sensing of Environment*, 73(2), 227-235.

Twigt, D., Rego, J. L., Tyrrell, D., and Troost, T. (2011). "Water quality forecasting systems: advanced warning of harmful events and dissemination of public alerts." *Proc., Proceedings of the 8th International ISCRAM Conference*, Lisbon, Portugal.

Utah DEQ. (2016). “Utah Lake, Jordan River, Canals Algal Bloom 2016.” <https://deq.utah.gov/legacy/divisions/water-quality/health-advisory/harmful-algal-blooms/bloom-events/bloom-2016/utah-lake-jordan-river/index.htm> (June 19, 2018)

CHAPTER 5

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

5.1 Enhancing and Supplementing the Water Quality Record through Remote Sensing

Remote sensing enhances both limited historical records of algal blooms and adds to ongoing monitoring of water quality. Several benefits demonstrated by the historical application in the GSL system include: improved spatial and temporal coverage, alternative measures of algal blooms, better historical context for evaluating past and ongoing conditions, and more holistic monitoring strategies.

The application of empirical remote sensing models to historical imagery in the GSL system produced a more continuous record with a greater frequency of estimates. It also extended the historical record for each of the lakes by at least a decade. This more complete and extensive record enables exploration of trends over the long-term and on a monthly scale, which was not previously possible given the sporadic nature of the field sampling record. Historical records of algae biomass were also enhanced by the additional spatial coverage enabled by applying the remote sensing models to the entire lakes. Instead of being limited to the locations used historically to monitor the lakes, the remote sensing models provide estimates for the entire lake, including areas that have never been monitored before. A collection of maps produced by the full-lake application over the historical period of 1984-2016 is provided for Utah Lake, GSL, and Farmington Bay in Appendix A, Appendix B, and Appendix C, respectively. The additional spatial coverage also enables alternative measures, including the average and maxima concentrations of chl- *a* over the entire lake, more complete estimate of variability in concentrations, and an estimate of the spatial extent of blooms.

An important benefit of having a more complete view of historical water quality conditions is improved context for current and future conditions. The relatively recent increased attention (especially via news and social media coverage) can give an impression of certain trends that are difficult to verify with a limited field record. For example, the 2016 Utah Lake bloom received widespread media attention, and the remotely sensed record indicates an increasing trend in extreme chl- *a* concentrations. However, it also indicates that large blooms of similar or greater magnitudes (based on chl- *a* and spatial extent) have occurred in previous years. The enhanced record provides valuable context in which to place current conditions; it is not enough to conclude that conditions are getting worse simply because the public hears about them more often. An improved historical context is critical for providing an accurate description of trends and evaluating whether conditions actually are getting worse, and what aspects of algal blooms are changing. Evaluating the historical context and long-term changes in algal blooms (rather than simply evaluating a single event or the current state), also helps identify conditions or factors that trigger or contribute to poor conditions. The utility of a long-term record was demonstrated in Chapter 3, as relationships to local weather and seasonal climate conditions were evaluated in a much more robust manner than the evaluation of similar variables for a single bloom event (Page et al. 2018). The long-term records of weather, climate, and bloom conditions provided a range of conditions, which leads to a more complete evaluation of how these factors contribute to algal blooms (rather than simply observing what conditions existed prior to and during a single event).

Ultimately, the alternative measures of algal biomass provided by remote sensing cannot completely replace other methods that may be more precise or offer better temporal resolution (e.g. grab samples that can be used to determine a broader range of constituents or automated buoy networks that provide real-time estimates); rather, they complement these other monitoring methods. The end result of having multiple sources is a more holistic and nuanced evaluation of algal biomass and blooms over time and over the spatial extent of the lake system.

5.2 Recommendations for Future Remote Sensing Model Development and Implementation in other Regions

Chapters 2 and 3 highlighted several important considerations that should guide the development and implementation of remote sensing models for estimating algal blooms: accounting for unique characteristics of the study area, and operating within constraints of data availability and the spatiotemporal relationships of remote sensing.

The study area of the Great Salt Lake, Farmington Bay, and Utah Lake is a system with unique characteristics that directly influenced the approach for historical remote sensing of algal blooms. This includes physical characteristics of the lakes: color and turbidity, shallow depths and bottom reflectance, etc. as well as characteristics of the monitoring records: frequency, extent, methods, and measures used by various monitoring agencies. The lake-specific approach used for this lake system was supported by the knowledge of these distinct characteristics and was further supported by the observations of temporal variability (presented in Chapter 2). The differences in known physical characteristics that affect optical properties and differences in observed variability patterns for this lake system suggest that generic empirical models may not be appropriate for large lake systems or even individual lakes where characteristics are highly variable.

In cases where a long-term history is of interest, the quality of satellite imagery and data products is limited by historically operational instruments and available datasets. Future instruments and datasets with improved spectral, spatial, and temporal resolutions that are better suited for HAB remote sensing can be used to improve recent and ongoing histories. However, in order to gain information about the more distant past, some concessions may be necessary in using sub-optimal, but available data. This places an important constraint on the accuracy of trend analysis; historical records of instruments with revisit times that are longer than bloom events may fail to capture critically poor conditions or even the occurrence of a bloom entirely. Similarly, the records could be missing periods of favorable water quality conditions. When field records are much more limited than the satellite records, it is especially difficult to know whether the remotely sensed record results in an underestimation or overestimation of conditions. At any rate, this uncertainty should be accounted for when using the trends to support evaluations of the lakes.

A similar constraint exists for field records that use sub-optimal measures for HABs. While historical records are available for chl- *a*, and this measure remains prevalent for many monitoring agencies, it fails to distinguish between species or provide information about toxins. As other measures of HABs (e.g. cell density, phycocyanin or toxin concentrations) become available, future and ongoing remote sensing models should adapt and use these new measures.

5.3 Exploration of Contributing Factors to Algal Blooms

Relationships to local weather, seasonal climate, and hydrology were explored on multiple scales in Chapters 3 and 4. Immediate effects of local weather events were varied throughout the lakes in the GSL system. This suggests that these short-term influences may not be generalizable for all lakes. Rather, observed influences may be used by local monitoring and management entities to identify specific locations that are of interest or merit additional attention after a weather event. Examples include the increase of concentrations at only two sites in Utah Lake immediately following rainfall events, or the difference in general responses to high air temperatures between sites in Utah Lake and GSL.

Study of the seasonal climate effects revealed that the relationships between these external variables and measures of algal biomass are highly complex. For example, winter snowpack conditions (measured via SWE) described significant differences in algal biomass in Utah Lake, but this effect is lagged by several months (the snowpack-driven runoff into the lake occurs long before algal blooms occur). Influence of temperature was also complex. At the short-term scale, there were increases in chl- *a* concentrations for some locations in Utah Lake immediately following high temperatures, while concentrations decreased for locations in GSL under the same conditions. Despite the lack of significant correlation between higher seasonal temperatures and annual chl- *a* concentrations, there were other indicators that temperatures are an influential factor. Over the historical time period, there were coincident increasing trends in seasonal temperatures and extreme chl- *a* concentrations, as well as an earlier shift in timing of chl- *a* extremes. In summary, the interactions between temperatures and algal biomass are complex and depend on the scale and lag between variables as well as the characteristics of the algae populations (which differ from lake to lake, even within this system of lakes). The complexity of this and other factors should be stressed when describing the causes and contributing factors to avoid over-generalizing effects.

5.4 Recommendations for Exploring other Factors and Long-Term Forecasting

To conclude this work, there are a number of suggested opportunities for extending and improving upon this research in the future. The first recommendation is to adapt the forecasting framework for long-term planning applications. The framework demonstrated in Chapter 4 is aimed at short-term monitoring and advisory decisions; however, long-term planning and management decisions could also benefit from exploring how projected predictive variables will affect algal bloom conditions. Each of the variables explored over the long-term record or used within the forecasting framework has the potential to change in the future. These changes may come from continuing development of the urban area changes in water management practices (affecting stormwater runoff and stream inflows), or changing climate conditions. For example, in the western United States, there is particular concern over changes in the frequency and intensity of rainfall events and drought (Wuebbles et al. 2014). These changes could have varying influences, depending on the scale and the lake. In Utah Lake and Farmington Bay, increases in rainfall intensity could result in a more prevalent short-term flushing effect (with lower chl- *a* concentrations near major inflows). On the other hand, if there are longer periods with low or minimal rainfall, this could increase the likelihood of greater chl- *a* concentrations in Utah Lake. Another concern for changing climate is differences in snowpack, shifts in snowmelt patterns, and earlier timing of peak streamflow (Barnett et al. 2004, Stewart et al. 2004). These kinds of future changes could lead to differences in the lagged effects between spring streamflow and water quality responses; in this case, the structure and variables of the BN model would need to be adapted to reflect these changes. For long-term forecasting, web services would not be necessary, and climate change projections could be used instead of observational data as inputs or evidence to the predictive model.

Another suggested extension of this work is to explore other factors that aid in predicting future conditions. In particular, anthropic processes related to inflows and nutrient loading would be helpful not only in forecasting for monitoring and advisory planning, but also for guiding management and regulatory strategies. Examples of decisions that could be informed by incorporating these processes include wastewater treatment operations, application of fertilizer, and irrigation practices. While these are important factors that need to be explored, limited data availability poses a significant obstacle. Water quality parameters from wastewater treatment facilities are often only published at coarse scales (generally, monthly averages or maxima are reported) or for few years, making it difficult to evaluate short-term or seasonal influences on blooms. Similarly, nonpoint sources of inflows and nutrients are not regularly monitored or gaged like the main stream inflows, which makes it difficult to quantify these inputs or represent their relationships to bloom conditions at any scale. Future exploration of these other factors may consider using models that simulate the water quality response of the lakes in this system to various point and nonpoint sources of inflow and nutrients. Stochastic simulations of lake water quality could then be used to define conditional probability relationships (Ames et al. 2005). Again, the forecasting framework could be adapted to use model outputs in addition to observational data as evidence for inferring future conditions.

A final recommendation is to integrate remotely sensed data and data from the automated buoy networks into the predictive model and forecasting framework. As imagery becomes available, automated processing and model application would be needed to produce updated estimates of algal bloom conditions, which would then serve as evidence in the predictive model. Development of web services would facilitate data retrieval and integration of measurements from the buoy network. Each of these improvements and opportunities for further research would help improve current understanding of the water quality in these lakes, and lead to better monitoring and management strategies in the future.

5.5 References

Ames, D. P., Neilson, B. T., Stevens, D. K., and Lall, U. (2005). "Using Bayesian networks to model watershed management decisions: an East Canyon Creek case study." *Journal of Hydroinformatics*, 7(4), 267-282.

Barnett, T., Malone, R., Pennell, W., Stammer, D., Semtner, B., & Washington, W. (2004). “The effects of climate change on water resources in the west: introduction and overview.” *Climatic Change*, 62(1-3), 1-11.

Page, B. P., Kumar, A., and Mishra, D. R. (2018). "A novel cross-satellite based assessment of the spatio-temporal development of a cyanobacterial harmful algal bloom." *International Journal of Applied Earth Observation and Geoinformation*, 66, 69-81.

Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2004). “Changes in snowmelt runoff timing in western North America under a ‘business as usual' climate change scenario.” *Climatic Change*, 62(1-3), 217-232.

Wuebbles, D., Meehl, G., Hayhoe, K., Karl, T.R., Kunkel, K., Santer, B., Wehner, M., Colle, B., Fischer, E.M., Fu, R. and Goodman, A. (2014). “CMIP5 climate model analyses: climate extremes in the United States.” *Bulletin of the American Meteorological Society*, 95(4), 571-583.

APPENDIX A

HISTORICAL REMOTELY SENSED CHLOROPHYLL A IN UTAH LAKE

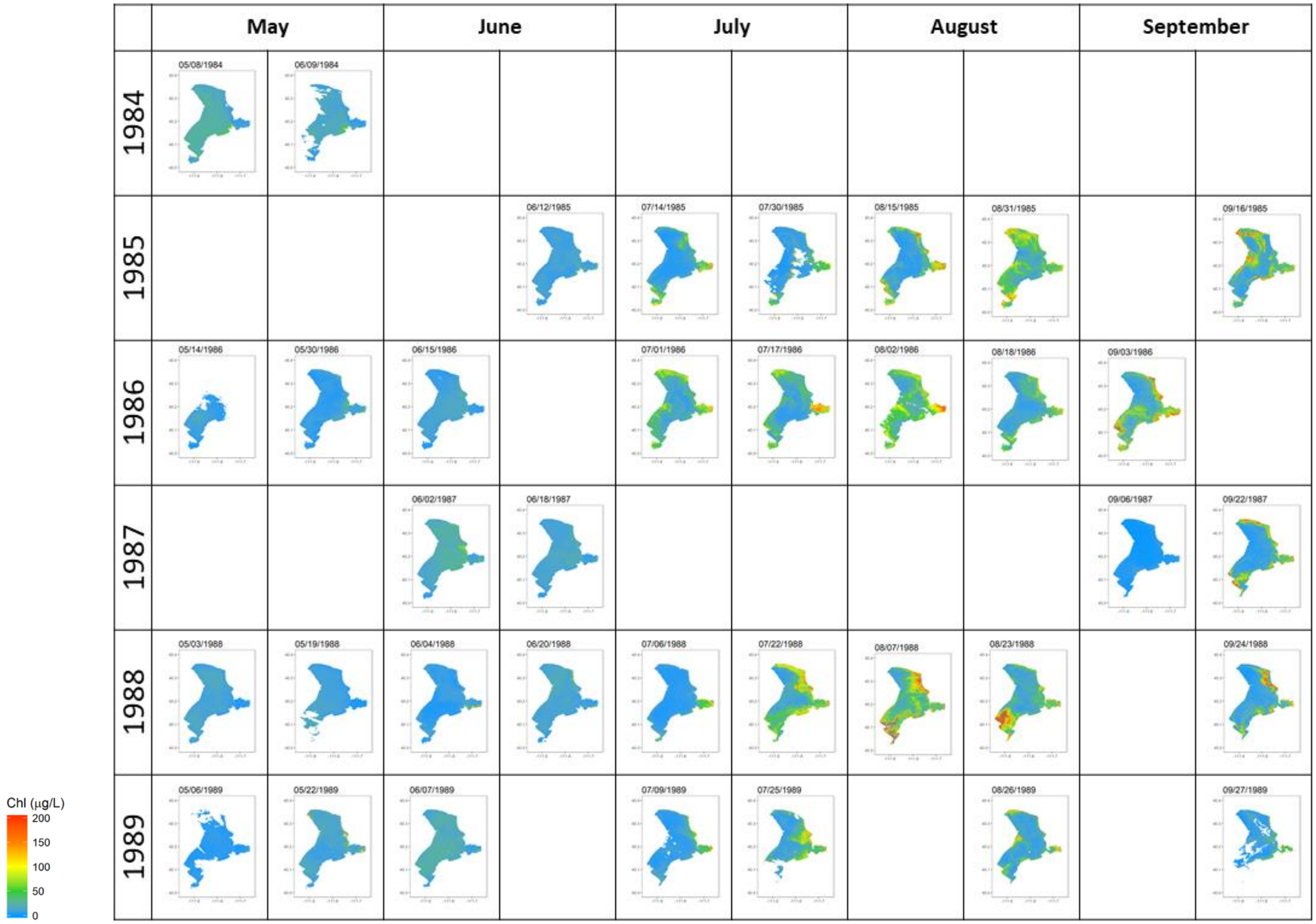

Figure A.1 Utah Lake Chl- *a* Concentrations from 1984-1989

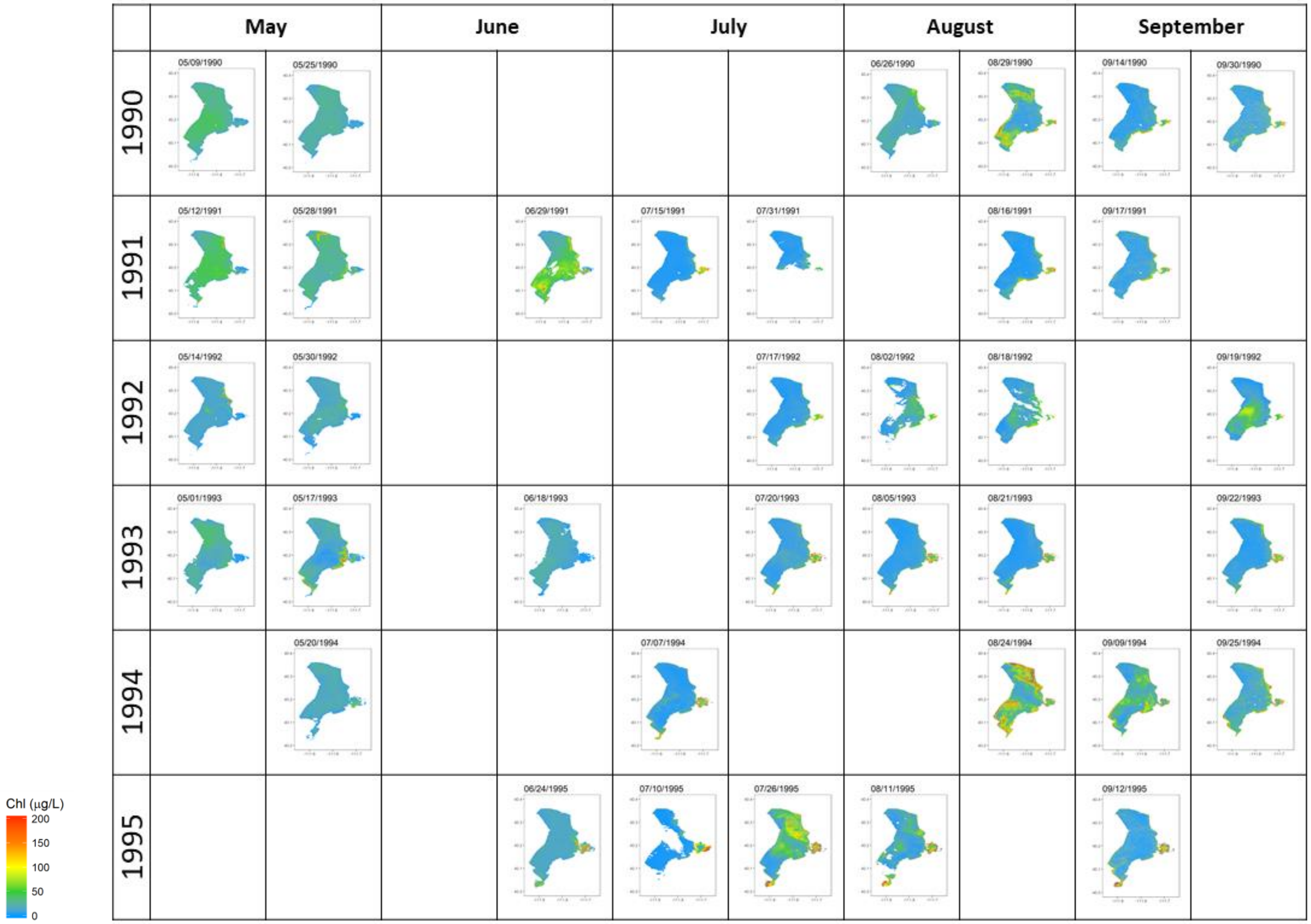

Figure A.2 Utah Lake Chl- *a* Concentrations from 1990-1995

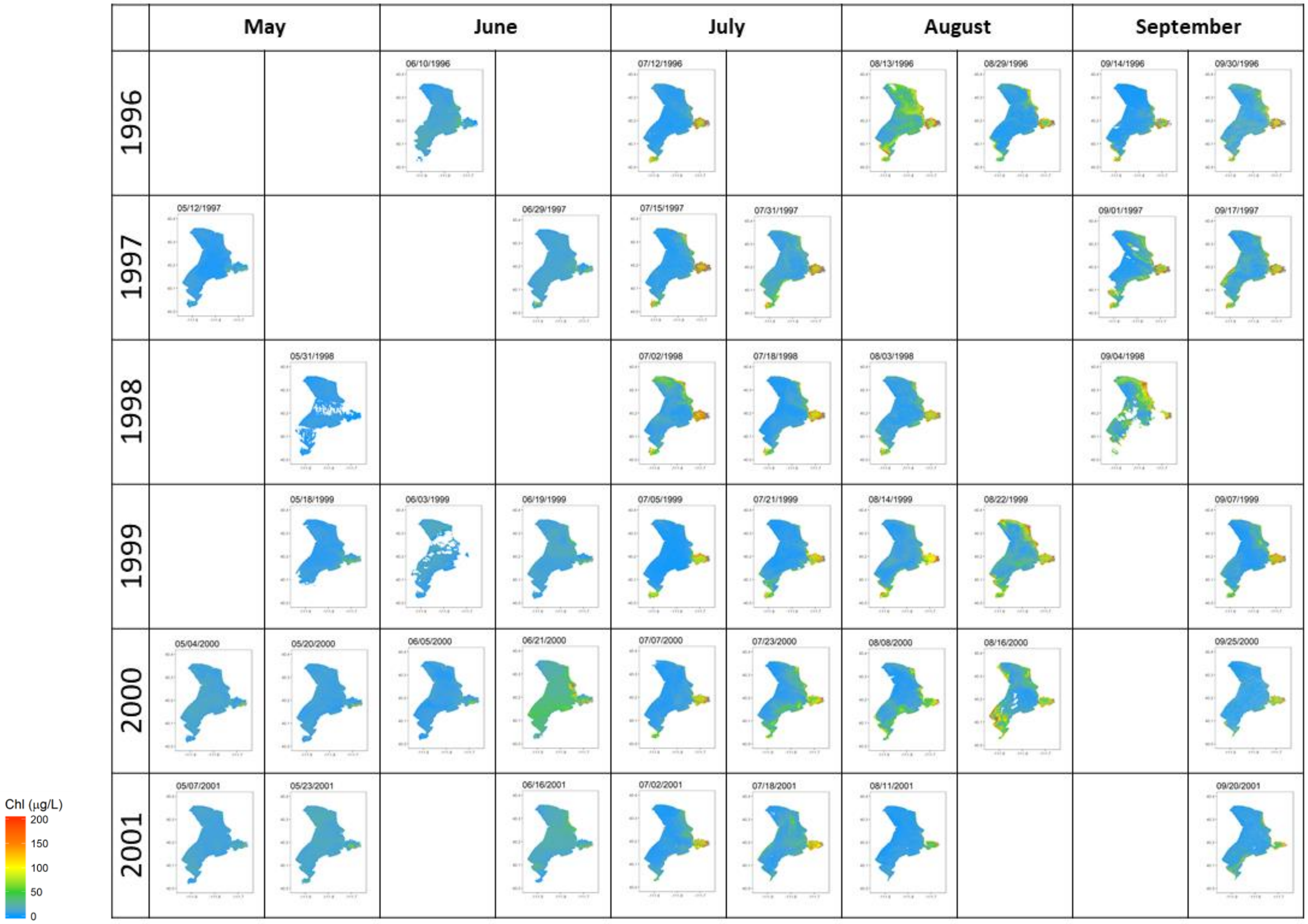

Figure A.3 Utah Lake Chl- *a* Concentrations from 1996-2001

Figure A.4 Utah Lake Chl- *a* Concentrations from 2002-2007

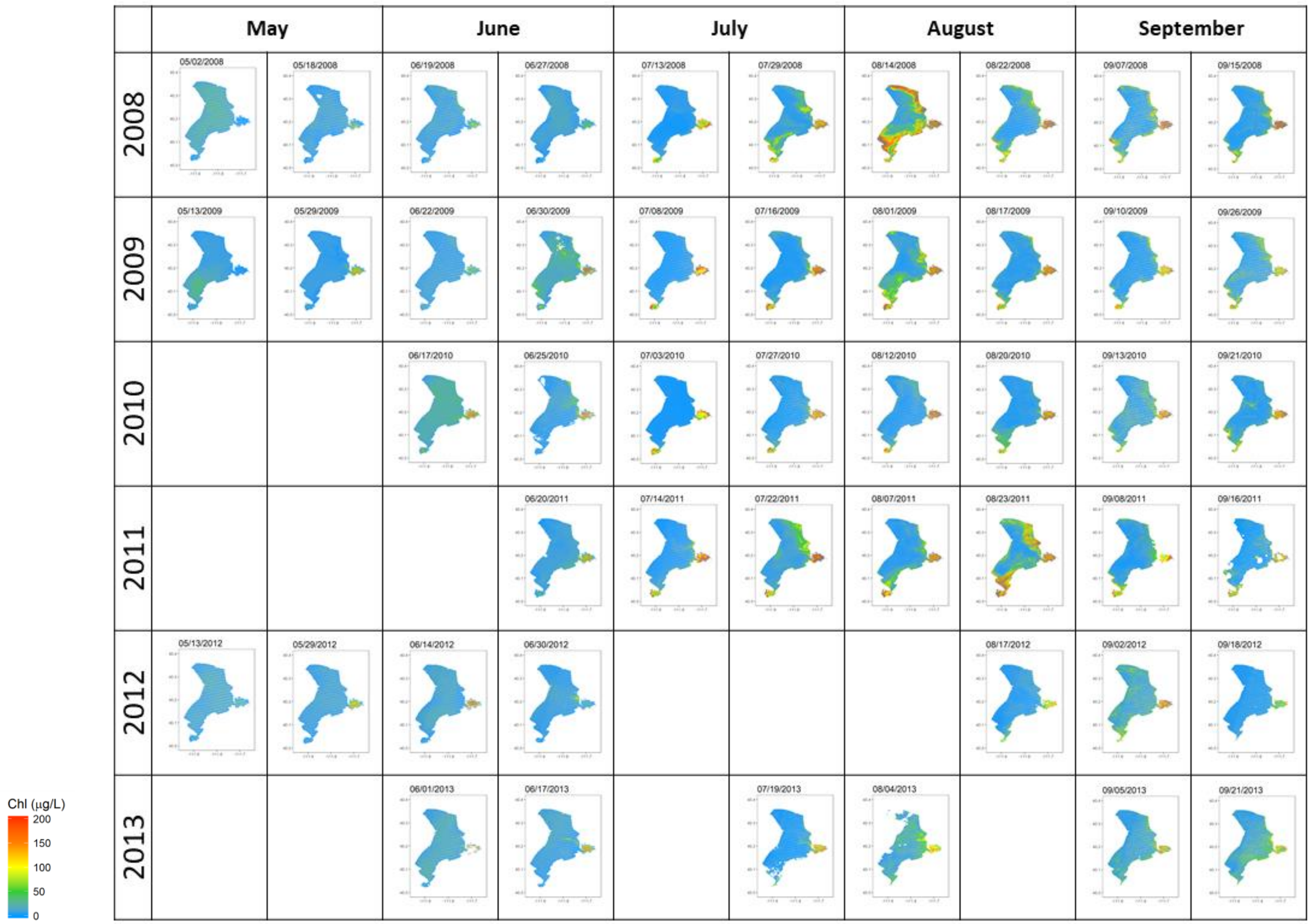

Figure A.5 Utah Lake Chl- *a* Concentrations from 2008-2013

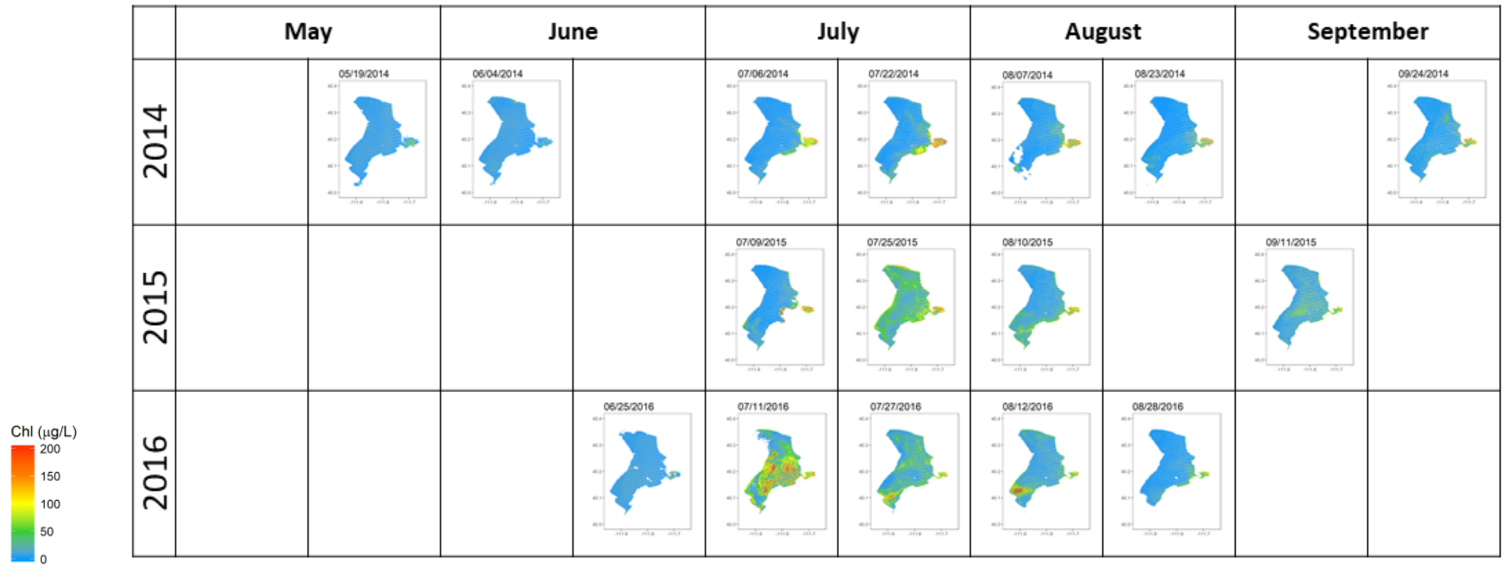

Figure A.6 Utah Lake Chl- *a* Concentrations from 2014-2016

APPENDIX B

HISTORICAL REMOTELY SENSED CHLOROPHYLL A IN GSL

Figure B.1 GSL Chl- *a* Concentrations from 1985-1989

Figure B.2 GSL Chl- *a* Concentrations from 1990-1995

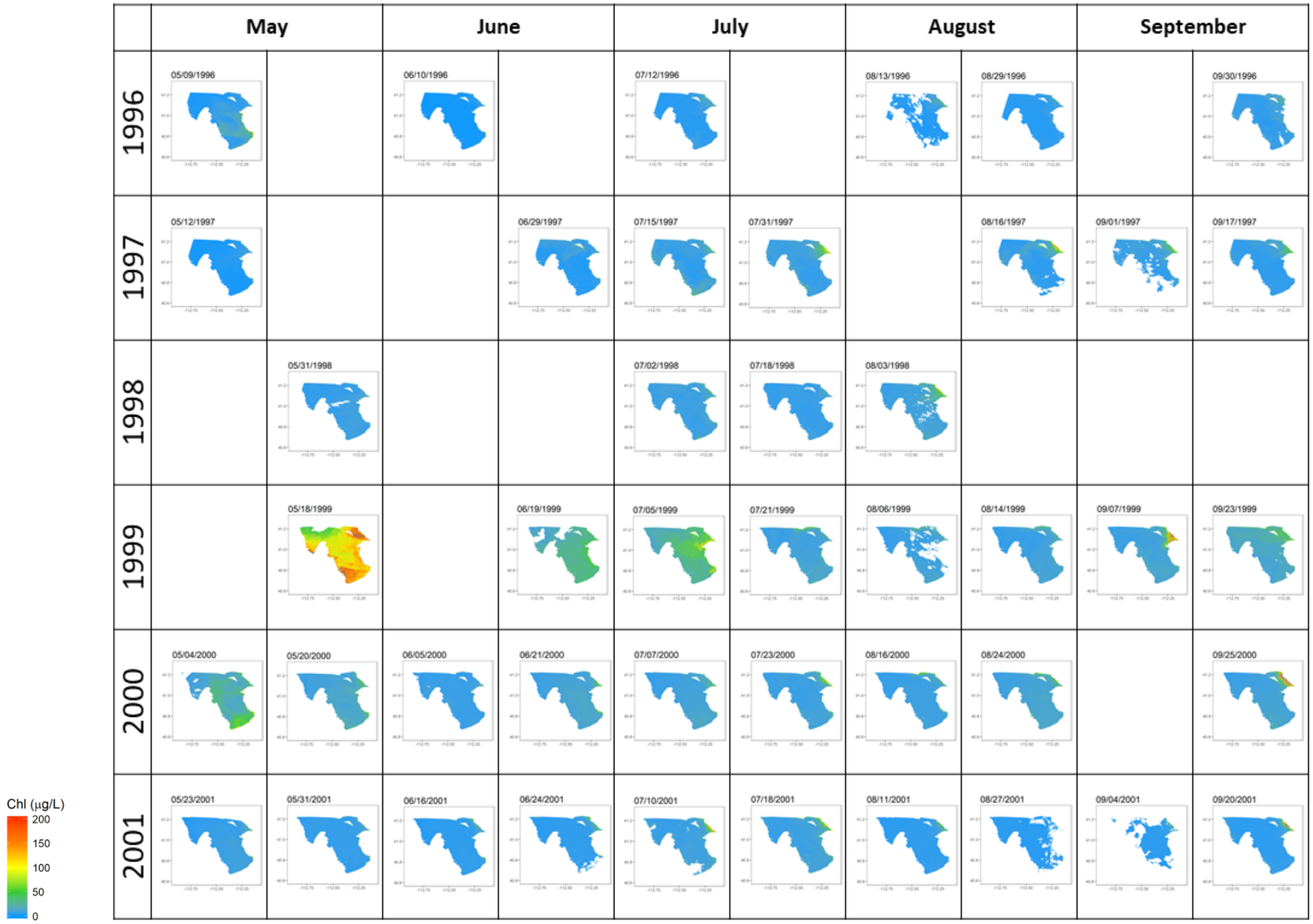

Figure B.3 GSL Chl- *a* Concentrations from 1996-2001

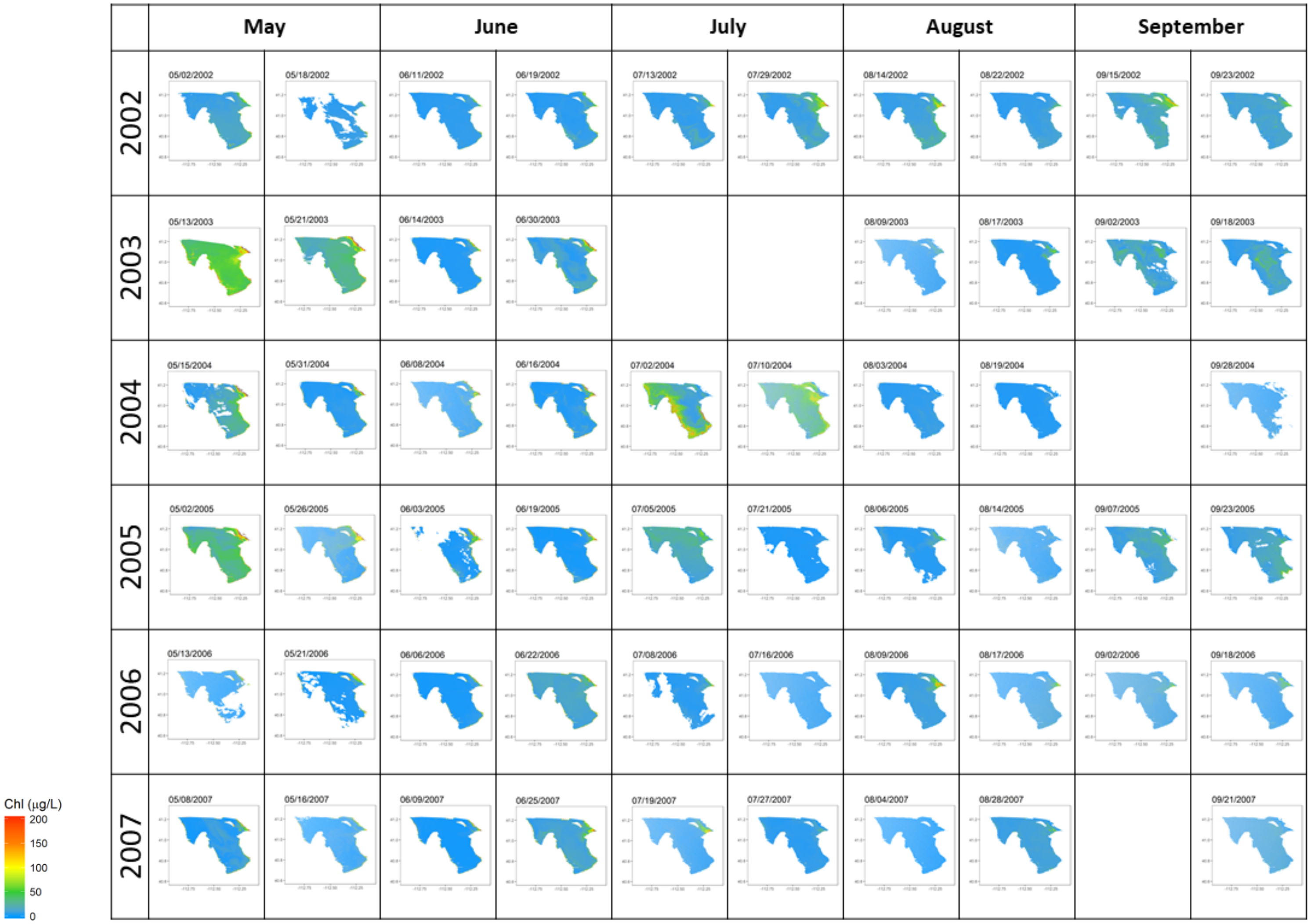

Figure B.4 GSL Chl- *a* Concentrations from 2002-2007

Figure B.5 GSL Chl- *a* Concentrations from 2008-2013

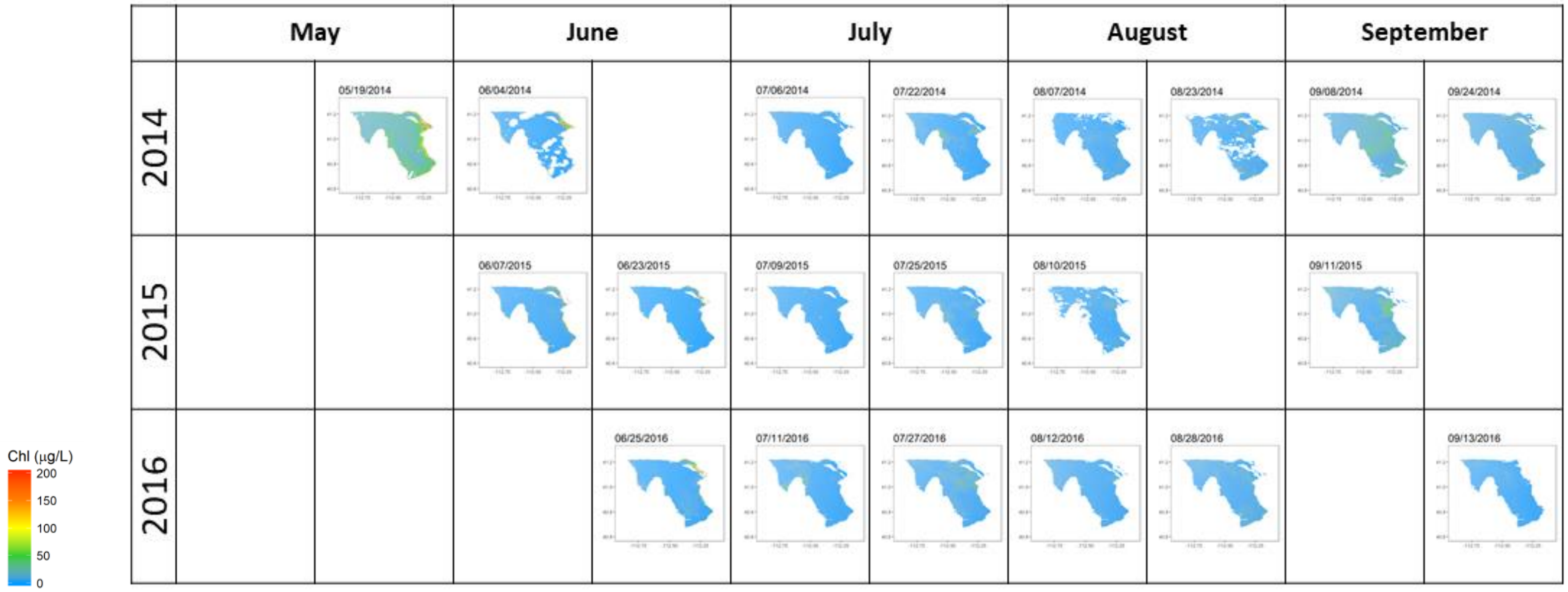

Figure B.6 GSL Chl- *a* Concentrations from 2014-2016

APPENDIX C

HISTORICAL REMOTELY SENSED CHLOROPHYLL A

IN FARMINGTON BAY

Figure C.1 Farmington Bay Chl- *a* Concentrations from 1984-1989

Figure C.2 Farmington Bay Chl- *a* Concentrations from 1990-1995

Figure C.3 Farmington Bay Chl- *a* Concentrations from 1996-2001

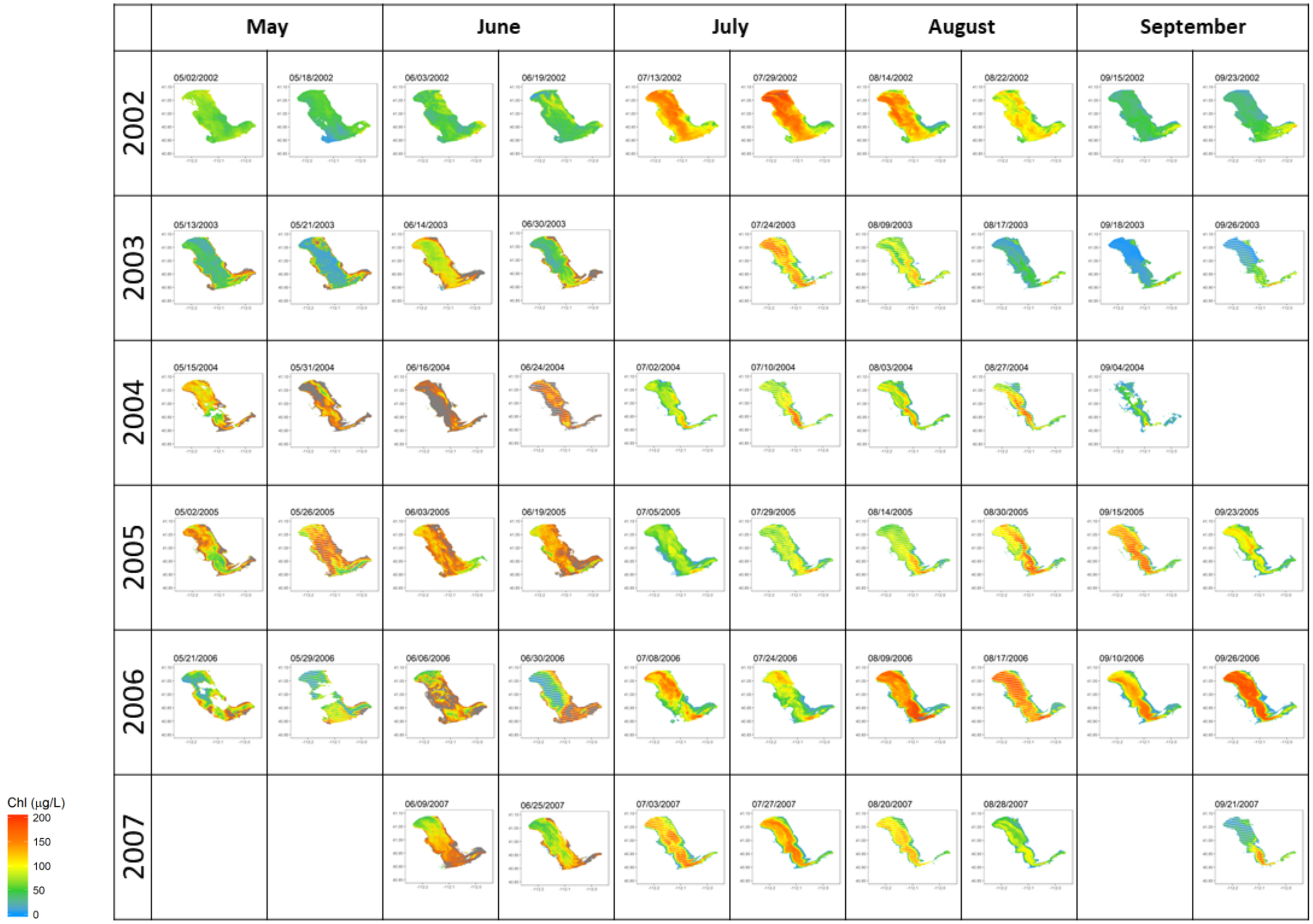

Figure C.4 Farmington Bay Chl- *a* Concentrations from 2002-2007

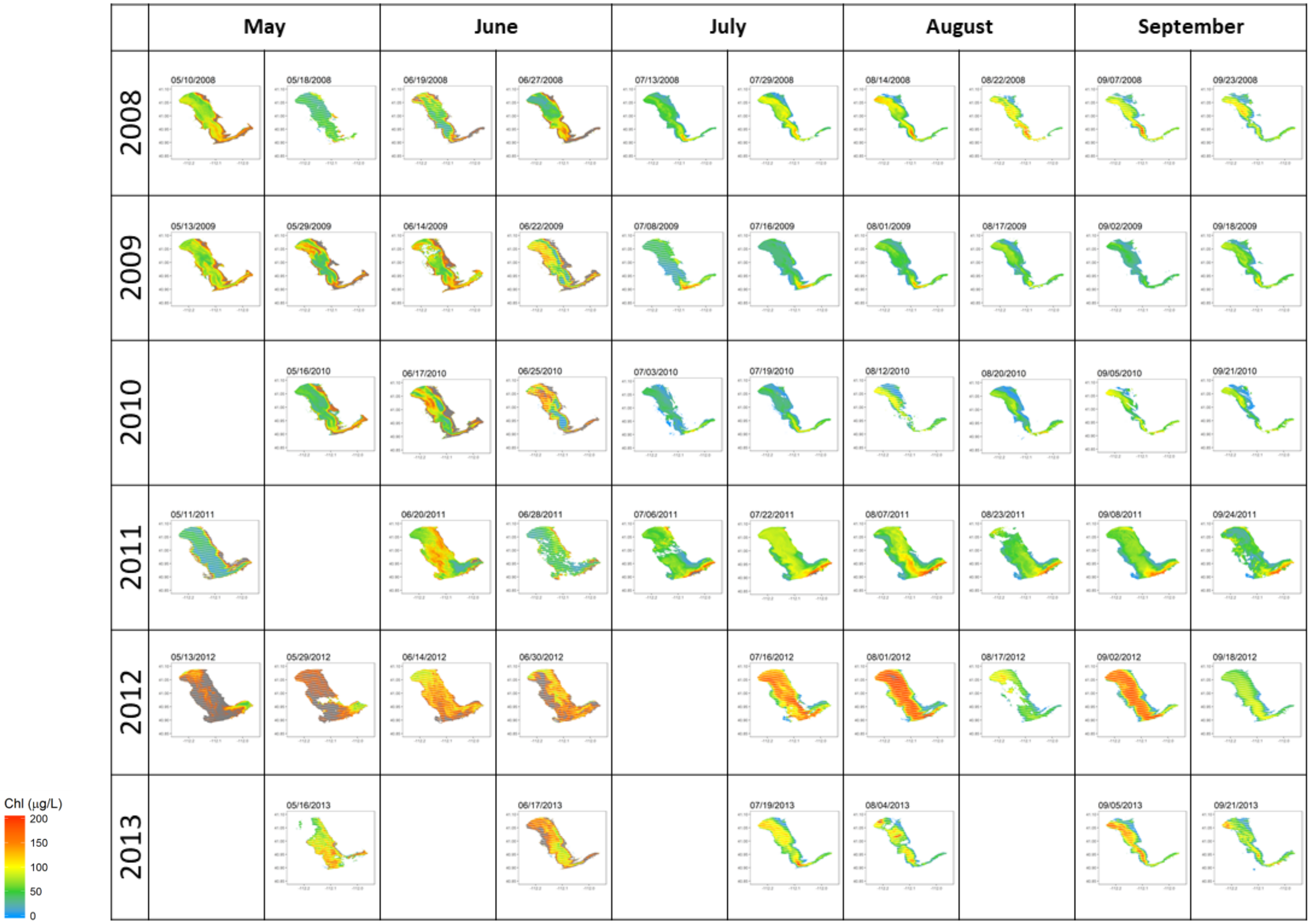

Figure C.5 Farmington Bay Chl- *a* Concentrations from 2008-2013

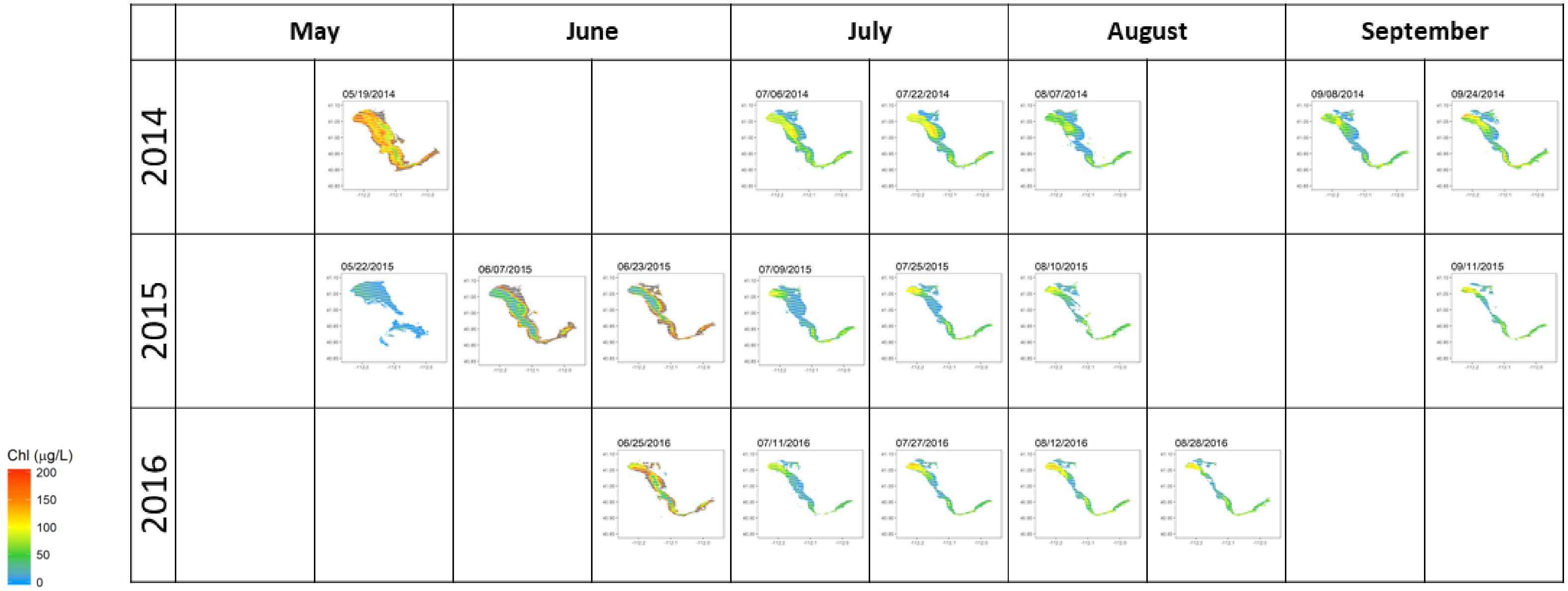

Figure C.6 Farmington Bay Chl- *a* Concentrations from 2014-2016

APPENDIX D

DATA RESOURCES

Collected data and scripts used in the analysis of this dissertation are published as resources and via repositories as detailed below:

- Data from the GSL repeated sampling analysis are available as a Hydroshare resource: https://www.hydroshare.org/resource/190caaeb411d4318a521f4410a3f9bbc/
- Reflectance data are available as a Hydroshare resource: https://www.hydroshare.org/resource/cbb0a0baacb74c57bb0b229362b0e7a1/
- The BN model for predicting seasonal extreme chl-*a* conditions is also published as a Hydroshare resource: https://www.hydroshare.org/resource/27f81cb47f814e32adc48f5fb9d02fa5/
- A public repository for the RSAlgaeR package and related scripts is hosted on GitHub: https://github.com/cahhansen/RSAlgae and via the CRAN repository: https://cran.r-project.org/web/packages/RSAlgaeR/index.html