Cyanobacterial harmful algal bloom modeling in eutrophic water bodies

by

Md Atiqul Islam

B.Sc., Bangladesh University of Engineering and Technology, 2016

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Carl and Melinda Helwig Department of
Biological and Agricultural Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Dr. Aleksey Y. Sheshukov

Copyright

Abstract

Harmful algal bloom (or HAB) is a global phenomenon in the rising trend of environmental concerns that impacts public health and the economy through declining water quality and toxicity. A rapid increase in cyanobacteria concentrations in water bodies is a primary cause of HABs. Enhanced eutrophication and warmer climate are considered vital driving factors for the proliferation of HAB events in the United States and worldwide. Dynamic modeling of cyanobacteria concentrations can help manage and reduce the impact of toxic blooms by better understanding the conditions for cyanobacteria growth and providing recommendations for early advisory warnings to the public for eutrophic water bodies in the agriculture dominated watersheds of the Midwest. In this study, sub-daily time series of cyanobacteria concentration and other environmental, physical-chemical variables were collected at the USGS sites in southcentral Kansas at Cheney Reservoir near the City of Wichita and in northeast Kansas at Kansas River near Wamego. Statistical analysis of the data revealed positive correlations between cyanobacteria concentration and water temperature, irradiation, phosphorus concentration, and storage volume. Correlation of dissolved oxygen depletion with cyanobacteria growth indicated an adverse impact of HABs on aquatic systems. A process-based mathematical framework for the kinetics of cyanobacteria growth was implemented at two sites considering bacteria natural growth, non-predatory loss, outflow washout, and accounting for the changes in water temperature (T), solar irradiance (I), and available nutrients (phosphorus [P] and nitrogen [N]). Four models were developed to facilitate examination of potential data limitation in sampling and continuous observations: (i) T-based, (ii) T, I-based, (iii) T, I, P- based, and (iv) complete four-factor model (T, I, P, N-based). The models were calibrated using continuous observations in 2013 - 2014 with time intervals from 2 days to 15 days (NSE = 0.41 to 0.71), and validated for 2018 (NSE = 0.56).

Simulations revealed model efficiency in short-term (one day to bi-weekly) forecasting of cyanobacteria concentration for both nutrient-rich sites. The performance of TIP-based and TIPN-based models was found acceptable for long-term forecasting in the Cheney Reservoir. Data sampling at a 15-day interval was found adequate for the forecasting of cyanobacteria growth. A stochastic modeling approach was applied to the TIPN model that converted a kinetic growth model to a modified Fokker-Planck equation for the probability density function of the cyanobacteria concentration to account for variability in influent nutrient concentrations and their impact on HABs. Several single storm event scenarios were simulated to evaluate the impact of high nutrient runoff into the lake on cyanobacteria. Stochastic model simulations showed that mechanistic modeling forecasting uncertainty increased along time propagation and higher uncertainty in initial concentrations of the cyanobacteria. The process-based mechanistic model was found to be useful for simulating future HAB events in the data-scarce eutrophic conditions, and preliminary insights into the stochastic modeling approach showed potential for future modeling direction under variable nutrient lake condition.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This report comprises the outcomes of my graduate research activities while studying at K-State. At this moment, I would like to thank everyone who helped and inspired me during this period. I would like to express my heartfelt gratitude to my major professor, Dr. Aleksey Sheshukov. Your invaluable guidance, supports, and inspirations made it possible to finalize this work. We had many discussions that motivated me to think deeper into the problems and triggered new questions. Thank you for investing your precious time.

I would also like to sincerely thank my supervisory committee Drs. Daniel Flippo and Trisha Moore. I am grateful for your keen interest in the topic. Your intuitive comments and discussions helped me to have an overall better understanding. Your enthusiastic participation inspired me to carry out my graduate studies.

Next, I would like to thank Dr. Daniel Tartakovsky from Stanford for his helpful suggestions to formulate innovative ideas. My special thanks to Dr. Jennifer Graham and Guy Foster from USGS for their advice on some critical points of the study. I acknowledge the Kansas Water Research Institute and USGS for funding this study project.

I also acknowledge the K-State community to facilitate smoothly continuing my graduate studies abroad. Finally, I want to thank my parents for their unconditional support and sacrifices for me. You are my driving forces for all of my accomplishments.

Dedication

Dedicated to my respectable parents for their everlasting love

Chapter 1 Introduction

1.1 Background

Cyanobacterial Harmful Algal Bloom (CyanoHAB) is one of the threatening global issues for many ecologically and economically essential lakes, rivers, and estuaries (Huisman et al., 2005; Hudnell, 2008). CyanoHAB is an intermittent but recurring water quality problem that creates taste-and-odor and toxin issues (Hudnell, 2008, p. 118). Toxic cyanobacteria bloom is an increasing worldwide phenomenon since historical times (Hudnell, 2008, pp. 108-111; Reynolds and Walsby, 1975, p. 438). Freshwaters, brackish, and seawaters blooms in countries of all continents and major oceans have been recorded (Codd, 1995, p. 150).

CyanoHABs have been reported in at least 48 states in the United States during 2010-2019 based on the Environmental Working Group (EWG, 2020) records, and at least 19 States had public health advisories alone in August 2016 (Graham et al. (2017, p. 2) cited USEPA, 2016). Cyanobacteria are considered the primary cause of harmful algal blooms (HAB) in Kansas (Kansas Department of Health and Environment [KDHE], 2020a). In 2019, the KDHE reported HAB contamination in 34 public lakes in 21 counties, which was the highest since the HAB monitoring program in 2010 (KDHE, 2020b). Periodic blooms in lakes, small ponds, and water bodies have the potential to develop toxin and taste-and-odor concerns that may cause significant economic, public health, and environmental problems. The United States alone was estimated $2.2 - $4.6 billion in economic costs annually due to blooms and eutrophication events (Dodds et al., 2009, p. 18).

“Eutrophication” is the nutrient enrichment of water bodies that triggers excessive biological productions and sometimes CyanoHAB (Fogg, 1969, p. 175). Eutrophic conditions are more susceptible to harmful blooms than low-level nutrient (oligotrophic) conditions (Dignum et

al., 2005, p. 66). During 1988-1993 survey periods ranging from 28 to 58 percent of the world lakes were eutrophic in Africa, South America, North America, Europe, and Asia Pacific regions (ILEC/Lake Biwa Research Institute, 1988-1993 cited in Bartram and Chorus, 1999, para 1.2). Anthropogenic activities, including agricultural operations, industrial works, and urbanization are linked with nutrient over-enrichment, causing more frequent bloom events (Pearl and Otten, 2013, p. 2). Data analysis by Glibert et al. (2010, Fig. 1) showed increasing global population trends along with energy consumption and fertilizers since 1960.

Anthropogenic alterations, nutrient (nitrogen and phosphorus) enrichment, light, temperature, water turbulence, turbidity, and biological interactions with other phytoplankton groups play vital roles in cyanobacteria growth (Fogg, 1969; Reynolds and Walsby, 1975; Pearl, 1988; Chorus and Bartram, 1999; Pearl and Otten, 2013). Cyanobacteria occurrences are expected to increase in a projected warmer climate and enhanced water eutrophication (Huisman et al., 2005, preface; Hudnell, 2008, p. 582 and p. 618). Climate change issues of rising temperatures, alterations in weather patterns, and stratification of aquatic systems can aid cyanobacterial growth, the dominance of harmful species over competitors, and their expansion worldwide (Pearl and Huisman, 2009). Cyanobacteria dominance period in eutrophic water bodies may vary depending on temperate or tropical regions. In temperate areas, cyanobacteria may dominate during summer, while it can be present all year round in tropical regions (Bartram and Chorus, 1999, sec1.2).

1.2 Problem Statements

Eutrophication of lakes and smaller water bodies is a severe problem in the Midwest. Eutrophication symptoms, including high turbidity, high pH levels, and low levels of dissolved oxygen, are prevailing in the Kansas lakes (KDHE, 2020c, p. 57). According to the National Water Quality Report to Congress (U.S. Environmental Protection Agency [EPA], 2017, p9), 55% of the

US assessed lakes are eutrophic or hypereutrophic. Even a higher percentage of reservoirs in the agriculturally dominated Midwest are eutrophic (KDHE, 2020c, p. 59). In Kansas, 207 of 322 surveyed reservoirs are experiencing nutrient over-enrichment. Currently, the KDHE uses three levels of weekly public advisory notification system based on microcystin toxic level and cyanobacteria concentration (KDHE, 2020a, p. 15), but lacks a HAB prediction model based on cyanobacteria dynamics.

Predictive tools integrating cyanobacteria dynamics, monitoring data, and remote sensing datasets are essential for short term forecasting of CyanoHABs and can be used to facilitate better algal bloom management (Ferguson, 1997, p. 2). Environmental, network, and ecosystem components can be essential parts of such predictive modeling tools (Janssen et al., 2019, p. 6). A mechanistic or process-based mathematical model relies on the mathematical description of physical processes and can integrate physical and biological factors causing the changes in cyanobacteria concentrations, including its growth and decay. Such a process-based mathematical model validated on actual HAB data can be used for short-term (days to seasons) to long-term (years to decades) forecasting of HABs and understanding the distribution, frequency, and intensity of cyanobacteria in waterbodies. A process-based model can be significant for longer-term HAB projections over statistical modeling approaches, which can be useful for very short-term forecasting (Ralston and Moore, 2020). The works of Canale and Auer (1982), Malve et al. (2007), and Chapra et al. (2017) are examples of process-based modeling based on phytoplankton and cyanobacteria dynamics. In addition, the validated model can be used as a tool for water-quality management in lakes, which can support public health advisory and developing early awareness (Guven and Howard, 2006a, p. 899; Glibert et al., 2010, pp. 266-271). Spatiotemporal variability of nutrients, sediment, and algae biomass in waterbodies can significantly affect

cyanobacteria growth and its incorporation in a model should be considered. These factors motivate researchers to develop and implement a process-based mathematical modeling approach for understanding the effects of environmental factors on bloom dynamics in a eutrophic aquatic system, especially for data-scarce lakes.

1.3 Objectives

In this study, we focus to understand the effects of various environmental variables on HAB dynamics based on literature and statistical analysis. The main goal of this study is to develop and implement a mechanistic modeling framework for cyanobacteria growth and HAB forecasting in two Kansas waterbodies (Cheney Reservoir and Kansas River), drinking water sources for the city of Wichita and northeastern Kansas (Ziegler et al., 2010, p. 4; Graham et al., 2012, p. 2). As part of the study, we also investigate and quantify the uncertainty of the developed model for understanding stochastic details of the forecasting approach. Available historical data sets will be used for model calibration, validation, and uncertainty quantification. Specific objectives are as follows:

1. Understand the impacts of environmental variables on cyanobacteria blooms.
2. Develop a mechanistic modeling framework for cyanobacteria forecasting considering system dynamics and major driving factors.
    a. Develop a process-based dynamic model combining physical growth factors (i.e., nutrients, temperature, irradiation) and biological interactions (growth and decay rate) of cyanobacteria based on historical in-situ data.
    b. Implement the developed model for short-term (day to a month) to long-term (months to years) forecasting of cyanobacteria growth.

c. Understand the applicability of the developed model in a data-scarce context by assessing conditions with different driving factors.

3. Quantify uncertainty of the developed mechanistic model in bloom forecasting.

This thesis consists of five chapters. Chapter 1 provides an overview of the study and presents background information, problem statement, and objectives. Chapter 2 reviews the literature on harmful algal blooms and underlying factors for cyanobacteria growth in water bodies. This chapter contains a theoretical background for understanding CyanoHAB and cyanobacteria dynamics and describes different algal growth models. In Chapter 3, the process-based mathematical modeling approach is presented following a statistical analysis to investigate correlation of cyanobacteria concentration with various environmental factors. Chapter 3 details the study area, explains the data collection procedure and discusses model development for HAB prediction. The methodology, results for selected study sites, and analysis are also presented in Chapter 3. Chapter 4 emphasizes developing and implementing stochastic modeling to understand the uncertainty of the developed mechanistic model in Chapter 3. Chapter 5 summarizes the findings from the study, discusses model shortcomings, and presents recommendations for future research.

Chapter 2 Literature Review

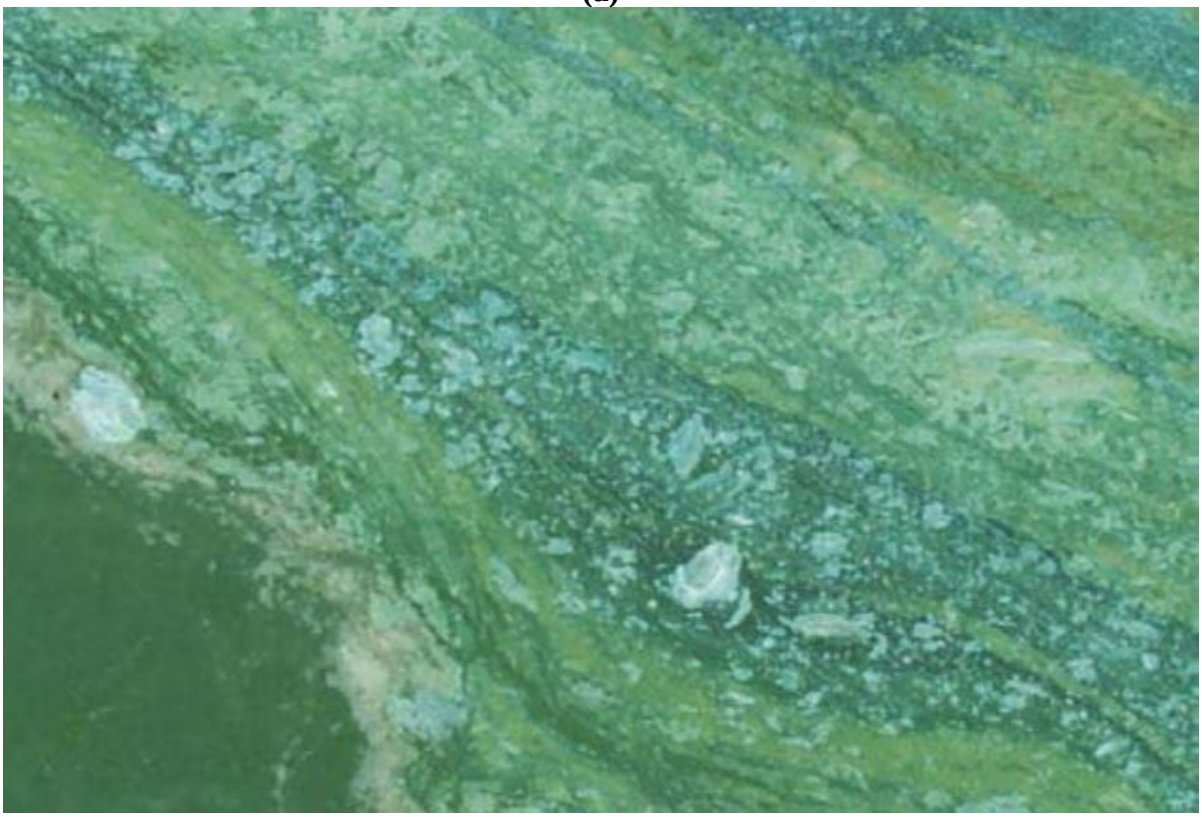
2.1 Harmful Algal Bloom (HAB)

2.1.1 CyanoHAB

Cyanobacterial Harmful Algal Bloom (or CyanoHAB) is the event of scum accumulation in a top layer of water due to excess growth of cyanobacteria in a favorable nutrient-rich environment (Reynolds and Walsby, 1975, p. 437; Graham et al., 2017, p. 5) (Figure 2.1). Harmful Algal Bloom (or HAB) happens due to the rapid growth and dominance of one or two phytoplankton species by efficient use of environmental resources (Whitton and Potts, 2002, p. 150). Harmful blooms are defined quantitatively as one or two nuisance species comprise 95-99% of the phytoplankton population in the range of 10,000 to 1000,000 cells/ml or more (Pearl, 1988, p. 825). Cyanobacteria and dinoflagellates are two primary HAB-forming species for freshwater and marine ecosystems respectively.

CyanoHABs impact public health by producing toxic materials that degrade reservoir water quality and require water treatment (Hudnell, 2008, p. 739). HABs can suppress the growth of other macrophytes by enhancing turbidity (Scheffer, 1997, p. 281) and deplete night-time oxygen resulting in lower fish production (Paerl and Huisman, 2009, p. 28). Cyanobacteria species produce peptides and alkaloids, thus making reservoirs not suitable for drinking, irrigation, fishing, and recreational purposes during HAB events (Codd, 1995, p. 152; Huisman et al., 2005, pp. 1-23). Cyanotoxins produced by cyanobacteria may cause severe animal and human health problems that can lead to death.

(a)



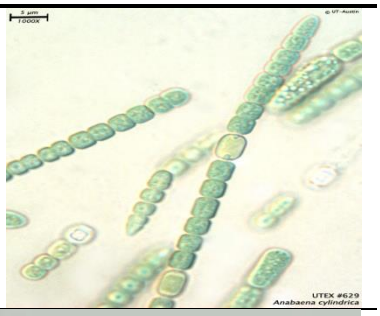
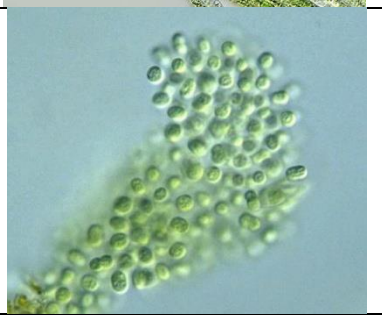
(b)

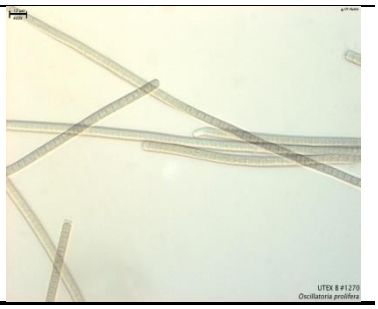
Figure 2.1 Cyanobacterial bloom in the (a) Ross Island Lagoon, Willamette River, Oregon (Courtesy Kurt Carpenter, USGS) and (b) Cheney Reservoir, Kansas (Courtesy KDHE).

2.1.2 Cyanobacteria Overview

According to fossil records, cyanobacteria are the oldest known prokaryotic structured oxygenic phototrophs bacteria (Whitton and Potts, 2002, pp. 13-35; Whitton, 2012, p. 33). Cyanobacteria contain photosynthetic pigment chlorophyll-a and function like the algal

Table 2-1 Major toxin producing cyanobacteria and their effects on humans (Pearl and Huisman, 2009. P. 29; New Hampshire Department of Environmental Services [NHDES], 2017). Pictures are taken from (Koning, 1994; Green Spanalga, 2009; Protist Information Server, n.d.).

Cyanobacteria Genera	Toxins	Side-Effects	Photographs
Anabaena	Anatoxins (neurotoxin), Microcystins (liver toxin), Saxitoxins (neurotoxin)	Nausea, vomiting, diarrhea, general malaise, severe thirst, skin and mucous membrane irritation, staggering, and paralysis	
Aphanizomenon	Saxitoxins (neurotoxin), Cylindrospermopsin (Liver toxin)	Numbness of lips and mouth extending throughout the body, motor weakness, respiratory and muscular paralysis	
Microcystis	Microcystins	Nausea, vomiting, diarrhea, general malaise, severe thirst, skin and mucous membrane irritation	

Oscillatoria	Anatoxins, Microcystins, Aplysiatoxins (dermatotoxin)	Nausea, vomiting, diarrhea, general malaise, skin and mucous membrane irritation	

community in the eco-aquatic system. They are also called blue-green algae (BGA) due to their bluish appearance from phycocyanin pigment. Commonly found toxin-producing cyanobacteria are nitrogen-fixing genera Anabaena, Aphanizomenon, Cylindrospermopsis, Lyngbya, Nodularia, Oscillatoria, and Trichodesmium; and the non- nitrogen fixers Microcystis and Planktothrix (Paerl & Otten, 2013). Description of major bloom-forming cyanobacteria and their effects on human health are described in Table 2-1. Bio-physiological properties of bloom-forming cyanobacteria contribute to their dominance in the aquatic system (Reynolds et al., 1987, p. 379). BGA can survive under a wide range of temperature, light, nutrition, saline, and alkaline conditions. Their special pigment enables efficient photosynthesis to survive in low irradiance. BGA can store excess nutrition that continues their growth during the nutrient-limited periods (Reynolds and Walsby, 1975). BGA can control their vertical distribution through the buoyant property, which helps absorb optimum light for bloom formation. This phenomenon helps them win over other competitors during nutrient and light insufficiency (Pearl, 1988, p. 825). Grazing effect on cyanobacteria is also minimal comparing with the other bacteria diminishing effects (Cerco and Cole, 1994).

2.1.3 Historical Blooms

Evidence of cyanobacterial blooms awareness dated back at least two thousand years ago in Europe and toxicity knowledge about scums among natives in North America, Africa, and

Australia (Huisman et al., 2005, p. 2). Indication of cyanobacteria presence in Llangorse Lake, Wales, can be found from the description of Geraldus Cambrensis in 1188 (Reynolds and Walsby, 1975, p. 438; Huisman et al., 2005, p. 2). Historical evidence on human illness, skin and respiratory irritation, gastrointestinal illness, and occasional deaths are documented as a result of bloom presence in drinking and recreational water (e.g., Huisman et al., 2005, pp. 1-13; Chorus and Bartram, 1999, ch. 3.4; Stewart et al., 2008, pp. 613-630). Official investigations (cited by Huisman et al., 2005, p. 2) by Hald (1833) in Denmark, Francis (1878) in Australia, and Benecke (1884) in Poland filed livestock, bird, and fish mortality from blooms are considered to be earliest-documented historical evidence. Fifty-two hemodialysis patients' death in Brazil from cyanobacteria toxin is one of the examples of human fatality (Huisman et al., 2005, p. 8). The earliest documentation of blooms in the US is from Minnesota. Evidence of gastrointestinal and other illnesses like nausea, vomiting, diarrhea, fever and eye, ear, and throat infections was documented in the US (Hudnell, 2008, p. 118).

CyanoHAB events have been increased worldwide since the 1960s, coinciding with increasing eutrophication (Hudnell, 2008, p. 113; Chorus and Bartram, 1999, para 1.3). In the US, cyanobacteria occurrences are also getting more documented than in the past (Hudnell, 2008, p. 113). The Environmental Working Group tracks the locations of HAB events across the US based on media reports (Figure 2.2). Monitoring media coverage from 2010 to 2019 shows that at least 48 states have experienced BGA blooms in this period (EWG, 2020). More than 50 locations in New York (186), Massachusetts (86), California (81), Kansas (65), Florida (59), and Washington (57) have recorded bloom occurrences during the last decade. At the same time, official records show that 50 counties in Kansas issued environmental advisories because of HABs (Figure 2.3).

Twenty-one Kansas counties issued HAB advisory for a period of 7 to 169 days in the year 2019 (Figure 2.4).

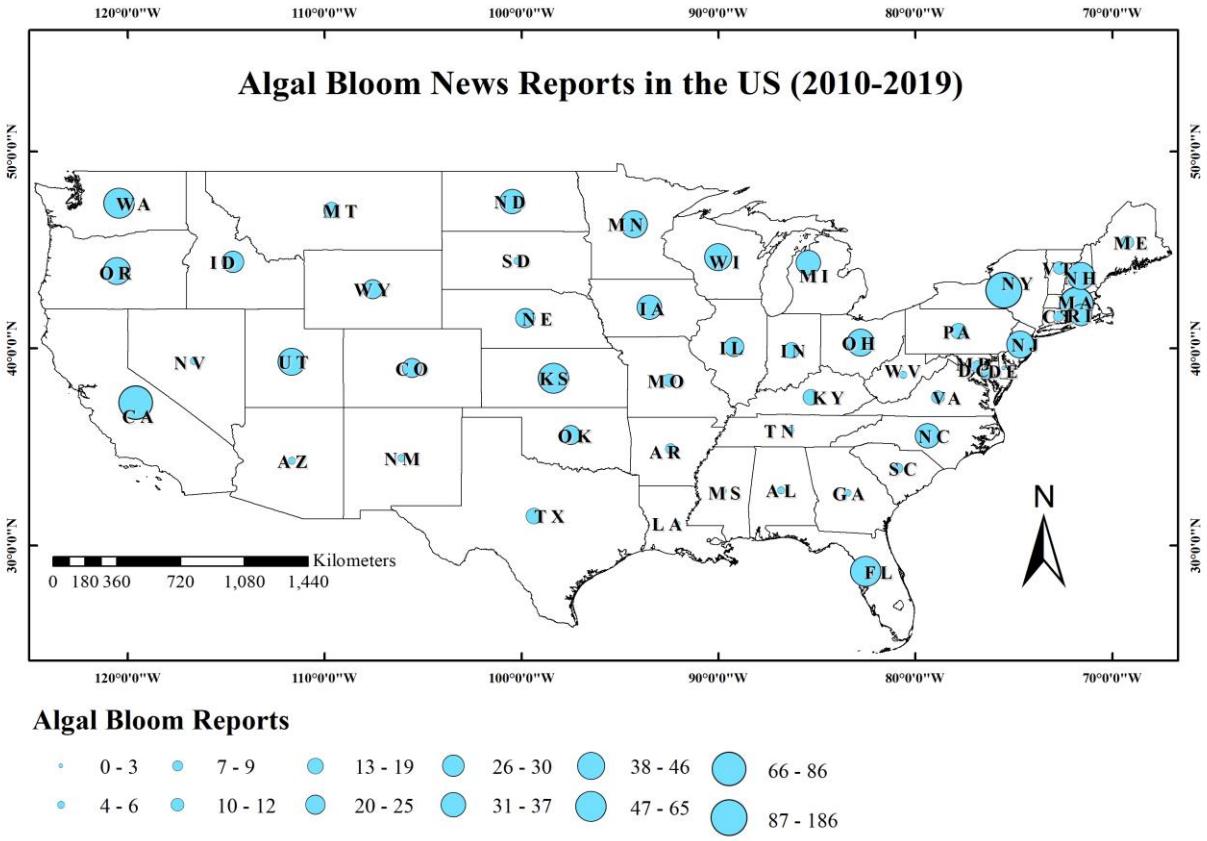


Figure 2.2 Harmful algal blooms in the United States during 2010-2019. At least 48 states have records of algal bloom based on media coverage. Map prepared based on collected data from EWG (2020) shows the number of reported bloom locations in different states.

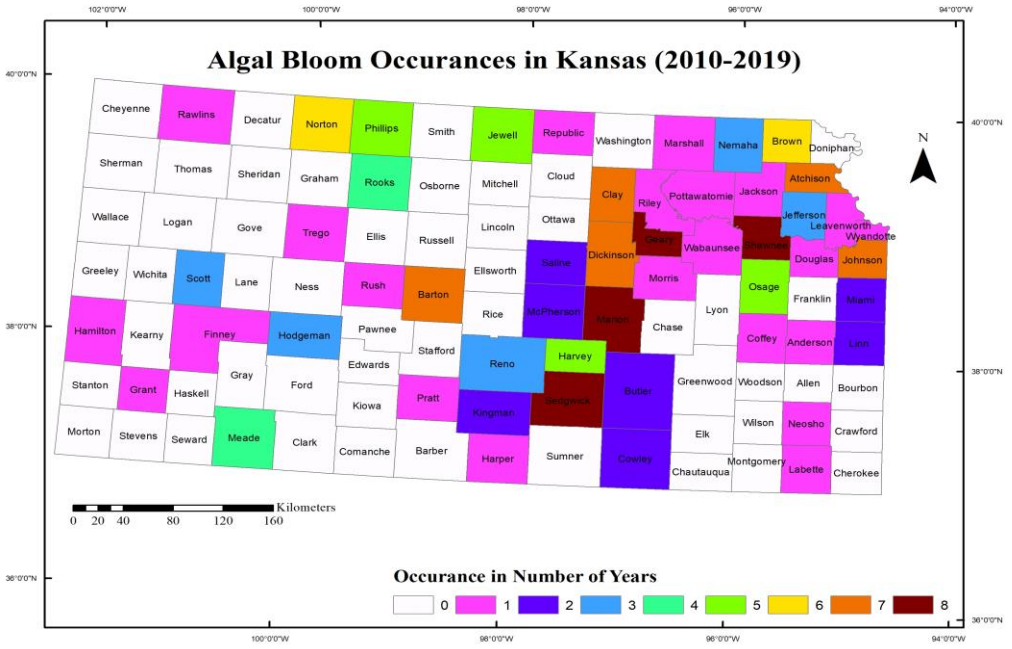


Figure 2.3 Advisory issued due to HABs in different county lakes and reservoirs in Kansas. Lakes and reservoirs in Sedgwick, Marion, Geary and Shawnee counties had been issued advisory in 8 different years during 2010-2019. Map has been prepared based on data compilation from KDHE (2020b)

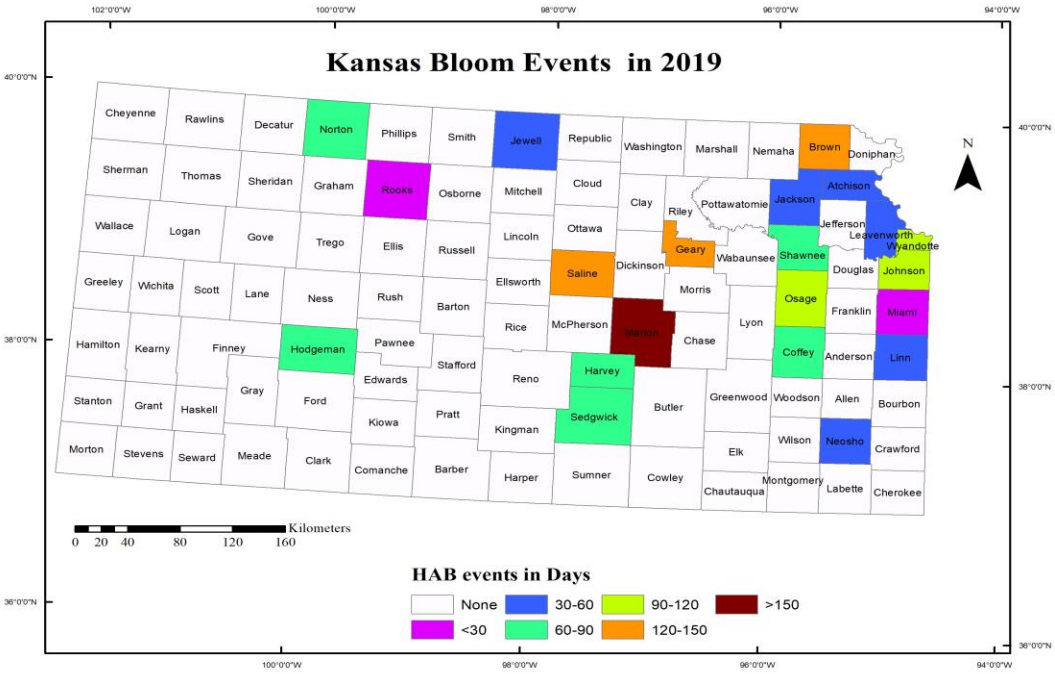


Figure 2.4 Map shows Kansas counties with harmful algal blooms and duration during 2019 (Based on data compilation from KDHE, 2020b)

Freshwater lakes like Lake Okeechobee and Ponchartrain in the US, Lake Erie and Lake Michigan in the US and Canada, Lake Victoria in Africa, Lake Taihu in China, estuarine and coastal waters like the Baltic Sea, the Caspian Sea, tributaries of Chesapeake Bay, North Carolina's Albemarle-Pamlico Sound, Florida Bay, the Swan River Estuary in Australia and the Patos are some examples of repetitive cyanobacterial bloom events (Pearl et al., 2011). Using remote sensing technique Ho et al. (2019) showed a significant increasing trend of global phytoplankton blooms in 68% of the studied lakes. Chapra et al. (2017) developed a modeling framework that projected HAB increase resulting from rising water temperature and nutrient-enriched conditions due to anthropogenic activities and changing watershed hydrology and water quality. The average duration of CyanoHAB occurrence is projected to increase to 18−39 days per year in 2090 compared to 7 days per year presently.

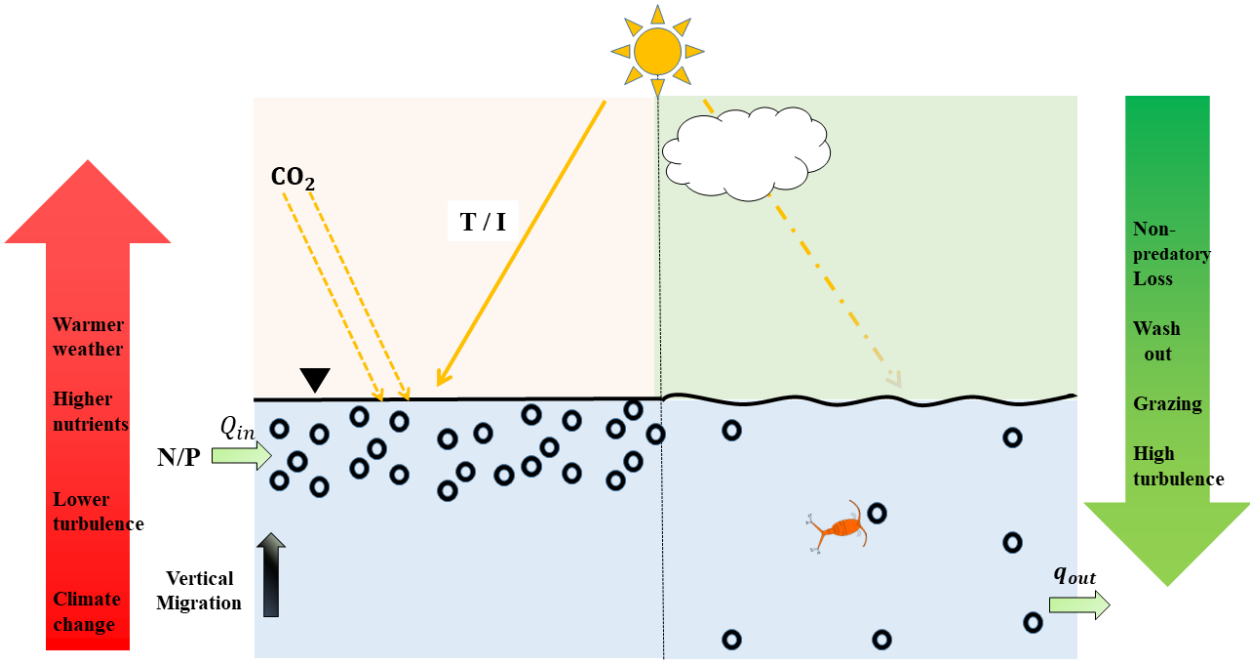


Figure 2.5 Schematic representation of major bloom driving environmental factors. The red arrow indicates factors favorable for cyanobacteria growth. The green arrow indicates factors that can reduce cyanobacteria proliferation.

2.2 Bloom Driving Environmental Factors

Significant factors that influence cyanobacteria growth are light, temperature, nutrient composition, vertical stratification, water retention time, surface water movement, turbidity, pH, and biotic interactions (Fogg, 1969; Reynolds and Walsby, 1975; Pearl, 1988; Chorus and Bartram, 1999; Pearl and Otten, 2013). Abundance and physiology of bloom-forming cyanobacteria and its seasonal change are crucial for the formation of HABs (Reynolds and Walsby, 1975; Whitton, 2012, pp. 155-187). Major controlling environmental factors for bloom formation are presented in the following sections and summarized in Figure 2.5.

2.2.1 Temperature Effect

Temperature controls cyanobacteria concentration by affecting the growth rate and increasing the stability of the water column that results in stratification (Visser et al., 2016). Reviewing the literature, Visser et al. (2016, pp. 151-152) suggest that the optimum temperature for cyanobacteria growth of 27-37℃ is higher than it is for most of its competitors except green algae. They also showed that the cyanobacteria growth rate increases faster than for green algae when the temperature rises. Thus, cyanobacteria can have a more favorable growth than other eukaryotic algae in warm weather. Pearl and Huisman (2009), citing Hutchinson (1957), stated that rising temperature decreases water viscosity and reduces resistance to vertical movement for cyanobacteria. Stratification in water triggered by both warm weather and vertical movement helps cyanobacteria to search for adequate light and nutrients. At the same time, non-mobile phytoplankton sinks to the bottom of the water bodies in these conditions. These jointly help cyanobacteria to create surface scum that sheds and works as a light obstruction to other competitors like diatoms and green algae (Reynolds and Walsby, 1975). As a result, rising air temperatures help to increase cyanobacteria abundance both directly and indirectly.

2.2.2 Light Effect

Light is required for photosynthesis and the growth of cyanobacteria. Cyanobacteria can maintain its growth at lower irradiance than its competitors (Reynolds and Walsby, 1975). The cyanobacteria growth rate increases along with increasing irradiance and reaches a specific optimum value when its cell density reaches the upper limit (Guven and Howard, 2006a, p. 900).

2.2.3 Nutrients Effect

Nitrogen, phosphorus, carbon, and silicon are essential nutrients required for aquatic organisms. Both inorganic and organic nutrients affect bloom formation (Pearl, 1988). Carbon and silicon are usually not limiting for cyanobacteria growth (Bowie et al., 1985), while nitrogen and phosphorus can limit cyanobacterial bloom (Reynolds and Walsby, 1975; Pearl and Fulton, 2006; Cerco and Cole, 1994). Reynolds and Walsby (1975) stated that phosphorus and nitrogen requirements for cyanobacteria are not significantly different compared to other competitors. Nutrient reservation capability of cyanobacteria during the nutrient-enriched period and subsequent use helps BGA to continue proliferation during periods when nutrients are scarce. Earlier studies suggested that phosphorus, a primary limiting nutrient, and low N-to-P ratios, are needed for cyanobacteria dominance. This conclusion was based on the assumption that nitrogen-fixing genera themselves produce an adequate amount of nitrogen required for cyanobacteria growth. However, Pearl and Otten (2013, pp. 3-4) discussed the importance of having both nitrogen and phosphorus for bloom because nitrogen-fixing genera can struggle to produce nitrogen due to insufficient micronutrients in it and the presence of other natural processes.

2.2.4 Climate Change Issues

A rise of CO_2, global warming, and shifts in seasonal precipitation patterns can promote harmful algal blooms (Visser et al., 2016). Since cyanobacteria can receive CO_2 directly from the

atmosphere due to its buoyant property, which is impossible for other phytoplankton groups (Pearl and Huisman, 2009), the enhanced atmospheric CO_2 can increase the photosynthesis process and encourage a bloom. Warming weather with prolonged periods of high temperatures and increased salinity in water bodies can help with water stratification and aid cyanobacteria abundances. Simultaneously, prolonged summer droughts and higher rainfall intensity can further worsen the situation by promoting calm water surface and higher nutrient loadings respectively. In conclusion, climate change can promote the occurrence of a suitable environment for cyanobacteria growth and increase CyanoHABs (Pearl et al., 2011; Pearl and Otten, 2013).

2.2.5 Other Bloom factors

Biotic and physical interactions along with nutrient enrichment play an essential role in bloom formation and persistence (Pearl, 1988). Biotic factors like bacterial interactions, grazing, and viral lysis can control bloom events (Pearl, 1988; Pearl and Otten, 2013). However, the zooplankton grazing effect on bloom control is not well explained (Pearl and Otten, 2013). The number of cyanobacteria, its toxicity, and ingestion rates creates issues for grazers. Furthermore, zooplankton grazing can control bloom formation by consuming non-cyanobacteria and hence indirectly helps harmful blooms. Viral lysis can also be significant for bloom decay. So, bloom formation, its sustainability, and collapse depend on complex interactions of various physical, environmental, and biological factors; and cannot be described as impacted by a single variable.

2.3 Cyanobacteria Bloom Models

Reynolds and Walsby (1975) indicated that temperature, light, and the medium's chemical composition controlled cyanobacteria growth factors. Smith (1985) modeled cyanobacteria as a combination of total phosphorus (TP), total nitrogen (TN), nutrition ratio (TN: TP), and mean depth (Z) of the lake. The study used multilinear regression analysis for determining a relationship

between cyanobacteria biomass and environmental variables (TN, TP, and Z). It was found that the regression relationship did not fit a nitrogen-fixing BGA well and high TN: TP ratio was contributing as a negative factor for the bloom. This model did not consider the effect of temperature, irradiation, and growth mechanism. Based on a review of past studies, Pearl and Otten (2013) discussed the TN: TP ratio concept and concluded that both total nitrogen (TN) and total phosphorus (TP) might be required for the bloom to form irrespective of N_2-fixing or non-N_2-fixing cyanobacteria genera.

Howard et al. (1995) developed a model that integrated light- and nutrient-limited bacteria movement and growth. They found that lake mixing influences vertical bacteria migration and affects cyanobacteria photosynthetic potential that results in reduced growth. Based on the Michaelis-Menten kinetics approach, they presented growth rate curves that combined light and nutrient conditions in the model. The findings revealed that the maximum growth happened at calm and eutrophic conditions consistent with other studies. That study did not include the effect of temperature on cyanobacteria growth.

Dalu and Wasserman (2018) studied potential drivers for cyanobacteria blooms using a collected dataset from a small reservoir that showed cyanobacteria abundance in the months of May-July. They found that the impacts of dissolved oxygen levels, water transparency, reactive phosphorus content, water depth, and chemical oxygen demand were significant on cyanobacteria growth using statistical analysis. They showed that improving water quality was able to mitigate the blooms with scenarios of climate change.

Havens et al. (2017) proposed a conceptual model including nitrogen, phosphorus along with zooplankton and hydrologic characteristics such as surface runoff as bloom regulating factors

in the presence of bloom-forming species. They indicated higher temperature, pH, and irradiance, decreased water flushing in a waterbody, and elevated nutrients as bloom-driving factors.

Harris and Graham (2016) found that data analysis with support vector machine, random forest, boosted tree, and cubist modeling techniques were able to predict cyanobacteria concentration while exploring 12 unique linear and non-linear regression models with 24 physical variables in the Cheney Reservoir in Kansas. Although Harris and Graham (2016) included dissolved oxygen and chlorophyll-a as modeling parameters, it is interesting to note that cyanobacteria use light energy by chlorophyll-a (Paerl & Otten, 2013) and chlorophyll itself could be the product of a BGA bloom. We emphasize that chlorophyll-a may be used as an indicator of the presence of cyanobacteria for HAB detection.

Harris et al. (2016) examined the effects of TN:TP and NO_3:NH_3 ratios on cyanobacterial secondary metabolites geosmin, methylisoborneol, and microcystin using data from four reservoirs in the Midwest US (Cheney Reservoir, Eagle Creek Reservoir, Geist Reservoir, and Morse Reservoir). It was revealed by the statistical methods, including principal component analysis, that geosmin, methylisoborneol, and microcystin were occurred primarily at low TN:TP ratio and higher cyanobacterial biovolume was evident at lower NO_3:NH_3 ratio.

Guven and Howard (2006b) reviewed deterministic mathematical models relevant to cyanobacteria growth and migration in the lakes. The review indicated that mechanistic models incorporate factors like temperature, nutrients, and light at a given time (Bannister, 1979; Laws Chalup, 1990; and Reynolds and Irish, 1997 in Guven and Howard (2006b)). They also discussed the importance of a process-based mathematical model over artificial neural network models. A mechanistic mathematical model incorporating system dynamics can better understand the blooming phenomenon impacted by underlying factors.

2.4 Process-based model functions

Phytoplankton dynamics are commonly expressed as a function of light, nutrients, and temperature (Bowie et al., 1985; Chapra, 1997; Canale and Auer, 1982; Malve et al., 2007). Nutrients considered in the model are nitrogen, phosphorus, carbon, and silica. Micronutrients like iron, molybdenum, and other trace metals for nitrogen fixation (Pearl and Otten, 2013) can be essential in oligotrophic conditions (Bowie et al., 1985), but generally are not included in the growth rate models and always considered available. Our study site, Cheney Reservoir, is a eutrophic lake (Stone et al., 2013) and so effect of micronutrients can be ignored during modeling. Table 2-2 presents a description of different functional relationships used for the phytoplankton growth rate model and can be relevant for the process-based cyanobacteria growth model. The growth model usually contains the maximum growth rate parameter at a reference temperature multiplied by the factors of temperature, nutrients, and light. The growth rate is maximum at optimum light and excess nutrients environment.

Table 2-2 Different functional models for φ_{Temp}, $\varphi_{\text{Nutrient, and}}$ φ_{Light}

Factor	Model form	Type	Applicability	References
φ_{Temp}	$\theta_{\mu}^{T-T_{Ref}}$	I (Theta model)	Phytoplankton modeled as a single state. Simulation performs lower than optimal temperature	Bowie et al.,1985 (p. 297); Chapra, 1997 (p. 605)
	$e^{-\kappa_1(T-T_{Ref})^2}$ for $T \leq T_{Ref}$ $e^{-\kappa_2(T_{Ref}-T)^2}$ for $T > T_{Ref}$	II (Optimum model)	When simulation performed on both side of optimal temperature and fits the species growth as asymmetric bell-shaped	Cerco and Cole, 1994 (pp. 4-11)
	0 for $T \leq T_{min}$	III	A simple linear form of Type II	Chapra, 1997 (p. 605)

	$\frac{T-T_{min}}{T_{opt}-T_{min}}$ for $T_{min} < T \leq T_{opt}$ $\frac{T_{max}-T}{T_{max}-T_{opt}}$ for $T > T_{opt}$	(Linear Optimum model)		
$\varphi_{Nutrient}$	$\prod_{i=1}^{m} \varphi_i$	I (multiplicative)	Multiple nutrients scarcity limits the growth rate	Bowie et al.,1985 (p. 307); Chapra, 1997(p. 608)
	$\min_{i=1,..,m} \varphi_i$	II (Liebig's law of minimum)	Modeling nutrient limiting behavior with the most scarce one	Bowie et al.,1985 (p. 307)
	$\frac{m}{\sum_{i=1}^{m} \frac{1}{\varphi_i}}$ Where, $\varphi_i = i/(K_i + i)$ $i = P, N, Si, C..$	III (Harmonic mean)	Considering the effect of in-between type I and II	Bowie et al.,1985 (p. 307, pp. 322-323)
φ_{Light}	$\frac{I}{K_I + I}$	I (Monod)	Saturation type response	Bowie et al.,1985; Chapra, 1997; Malve et al., 2007
	$\frac{I}{I_s} e^{(-\frac{I}{I_s}+1)}$	II	Optimum light concept	Steele (1965)
	$\frac{1}{K_e H} \ln \frac{K_I + 0.47I}{K_I + 0.47Ie^{-K_e H}}$	III	Considering the effect of depth from the water surface	Bowie et al.,1985 (p. 317)

2.4.1 Temperature factor, φ_{Temp}

Phytoplankton growth rate generally increases up to a certain optimum temperature and then decreases (Pearl et al., 2011; Pearl and Huisman, 2009; Canale and Vogel, 1974). Three temperature models are presented below (Bowie et al., 1985; Chapra 1997).

Exponential model:

The exponential model is used for the conditions when the temperature is below the optimum temperature, and phytoplankton is modeled as a single state variable in the system. Many past studies used the theta model (Type I of φ_{Temp} in Table 2-2) based on Arrhenius or van't Hoff

equation to consider the effect of temperature (Bowie et al., 1985). 20℃ is often used as a reference temperature for the exponential model (Compilation from Bowie et al., 1985).

Optimum model:

Cerco and Cole (1994) used Gaussian or bell shape function for temperature-dependent bacteria growth (Type II of φ_{Temp}). This model uses the concept of optimum temperature and can help to simulate various species at a time.

Linear model:

The linear relationship between growth rate and the temperature is defined as a linear form. The minimum value of the temperature in this model can be set when growth is zero. Another version of the linear model (Type III model) considers different linear slopes for temperatures below and above an optimum temperature for growth. This Type III model of $\varphi_{Temp}$is applicable for modeling phytoplankton growth in a broader range of temperature variations.

2.4.2 Nutrient factor, $\varphi_{\text{Nutrient}}$

The effect of nutrient-limited growth is simulated using a well-known Monod kinetics model (Monod, 1949) that assumes the first-order kinetics for phytoplankton concentration lower than half-saturation coefficient whereas zero-order kinetics is used for higher values. Different combinations for multiple nutrient limiting factors (multiplicative, minimum, or harmonic mean) are listed in Table 2-2. Type I model is applicable when more than one nutrient limits the growth rate. Type II model adopts the growth rate based on Liebig's law of the minimum. This assumes that the most unavailable nutrients, based on the half-saturation coefficient, control the growth. Type III model is valid when the concentration of the nutrients is not low (Walker, 1983 cited in Chapra, 1997).

2.4.3 Light factor, φLight

The light dependency of the phytoplankton can be modeled with a Monod formulation (Type I of φ_{Light} in Table 2-2). An alternative is the Type II model that accounts for optimum light conditions. Type III model is suitable for capturing the effect of light attenuation with an increase in depth from the water surface and incorporating the effect of turbidity on the light availability of phytoplankton.

Chapter 3 Harmful Algal Bloom: Mechanistic Modeling

3.1 Study Area

3.1.1 Cheney Reservoir

Cheney Reservoir (97° 50'16.11" W, 37° 45'32.99" N) is located in south-central Kansas near the city of Wichita at the east end of the North Fork Ninnescah Watershed (Figure 3.1). The reservoir was constructed in 1965 by the Bureau of Reclamation, the US Department of the Interior, for flood control and as a water supply source for the city of Wichita (Stone et al., 2013, pp. 1-2). Cheney Reservoir provides up to 70% of the water supply for the city of Wichita and more than 350,000 people (Ziegler et al., 2010, p. 5; *Cheney Lake Watershed Restoration*, 2011,

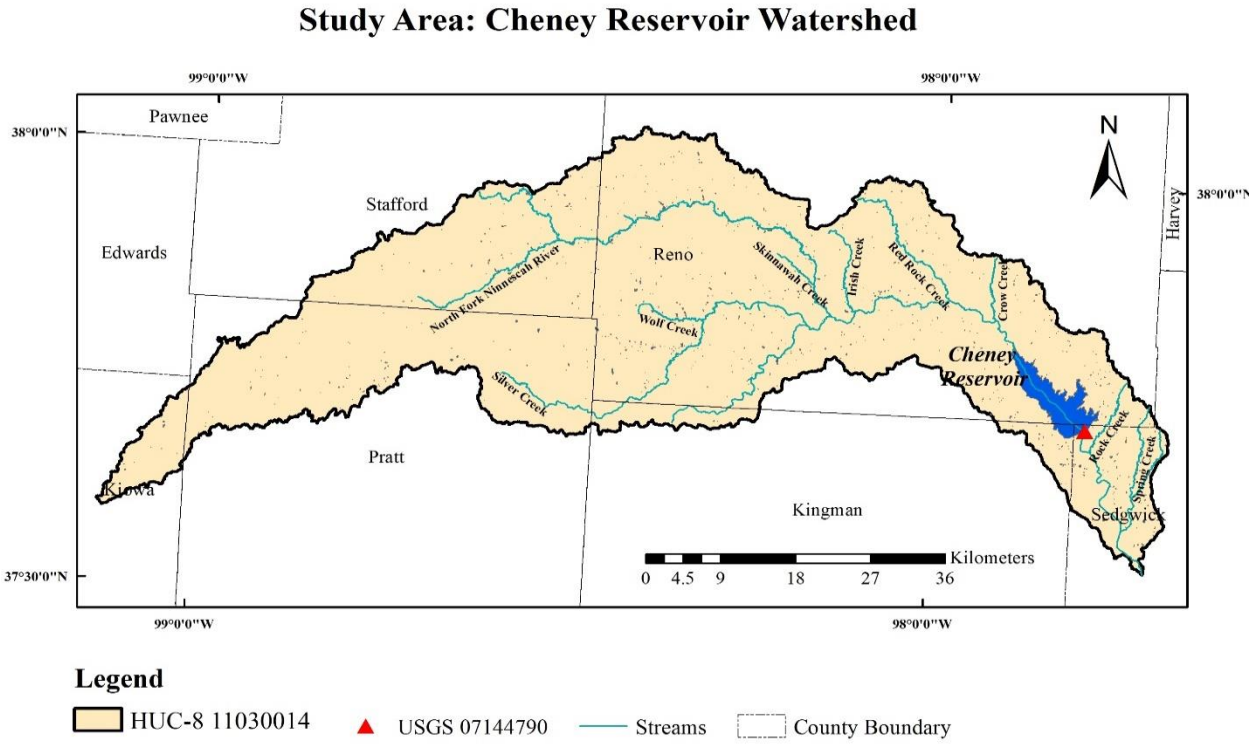


Figure 3.1 Cheney Reservoir and the North Fork Ninnescah Watershed (HUC-8 11030014). Red triangle shows a continuous water quality monitoring station near the dam (USGS 07144790).

p. 23). The reservoir is a shallow, eutrophic (average total phosphorus 105 µg/L) and well-mixed impoundment that is subjected to occasional thermal and chemical stratification (Stone et al., 2013, p. 2; Wang et al. 2003). The draining watershed for Cheney Reservoir is HUC-8 11030014 North Fork Ninnescah Watershed (2,415 km^2) with North Fork Ninnescah River as the main water stream. Other streams contributing to the watershed system are Silver Creek, Wolf Creek, Skinnawah Creek, Irish Creek, Red Rock Creek, and Crow Creek. Land use/land cover in the watershed is composed of cropland (58%) dominant, followed by rangeland (25%), and land in Conservation Reserve Program, or CRP (17%). Major crops are winter wheat (26%), corn (9%), soybean (4%), and grain sorghum (4%).

Cheney Reservoir has experienced historical algal bloom events since the 1990s (Smith et al., 2002) that caused taste-and-odor and toxic problems and increased water treatment costs (*Cheney Lake Watershed Restoration*, 2011, p. 16). Toxin-producing cyanobacteria Microcystis, Anabaena, and Aphanizomenon are commonly present in the reservoir (Graham and Harris, 2016). Historical algal bloom events with cyanobacteria cell counts of over 10,000 cells/ml in Cheney Reservoir from 2011-2020 reveals that cyanobacteria play significant roles for those events (Figure 3.2). Most events had cyanobacteria exceeding 20,000 cells/ml which accounted for 50%-100% of the total algal community. Cyanobacteria-related bloom events in Cheney Reservoir frequently exceeded the thresholds and created alerts and guidance levels that restricted the reservoir use as a drinking and bathing water source according to the World Health Organization (WHO) guidelines (Table 3-1, Chorus and Bartram, 1999, ch. 6.4). It can be noted that cyanobacteria alert levels 1 and 2 indicate 'moderate-high' to 'very high' risk categories for human health concerns respectively.

Table 3-1 Cyanobacteria management guidelines for drinking and bathing waters by WHO (adopted from Chorus and Bartram, 1999, ch. 6.4)

Cyanobacteria Concentration (#/ml)	Drinking Waters Alert Level	Bathing Waters Guidance Level
>200	Vigilance Level	
2,000	Alert Level 1	
20,000		Guidance Level 1
100,000	Alert Level 2	Guidance Level 2

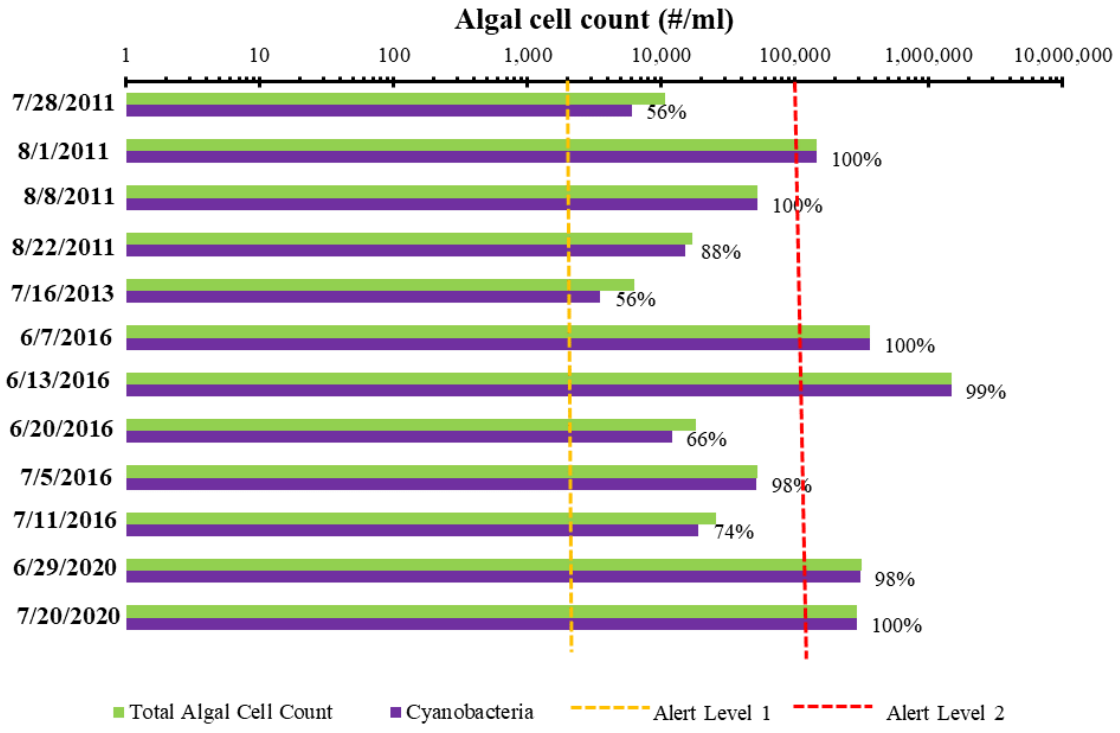


Figure 3.2 Major algal bloom events that exceeded 10,000 cells/ml in algal concentration in Cheney Reservoir since 2011. Green bars indicate total algal counts, and purple ones indicate cyanobacteria cell concentrations. Numbers in percentage indicate cyanobacteria concentrations in the total algal cell counts. Data based on KDHE (2020b).

3.1.2 Kansas River

Kansas River, located in northeastern Kansas, is a primary drinking water source for 800,000 people (Graham et al., 2012, pp. 1-5). Kansas River at Wamego (96° 18'19" W, 39° 11'54" N) has a drainage area of 143,174 km^2 and is situated at the lower Kansas River Basin (Figure 3.3). About 77 percent of the drainage area is agricultural land. Milford and Tuttle Creek controls water

flowing through the Kansas River, Wamego, and its water quality. Graham et al. (2012), citing Carney (2012) and communications from Neff (2012), stated no reported cyanobacteria blooms in the upper part of the Kansas River and occasional bloom events in the lower river basin. However, Milford Lake has frequent blooms, and water releases from that lake can be associated with cyanobacteria transport and taste-and-odor issues in the river downstream. Toxin producing cyanobacterial genera Anabaena, Anabaenopsis, Aphanizomenon, Aphanocapsa, Cylindrospermopsis, Microcystis, Phormidium, and Pseudanabaena were present at this site.

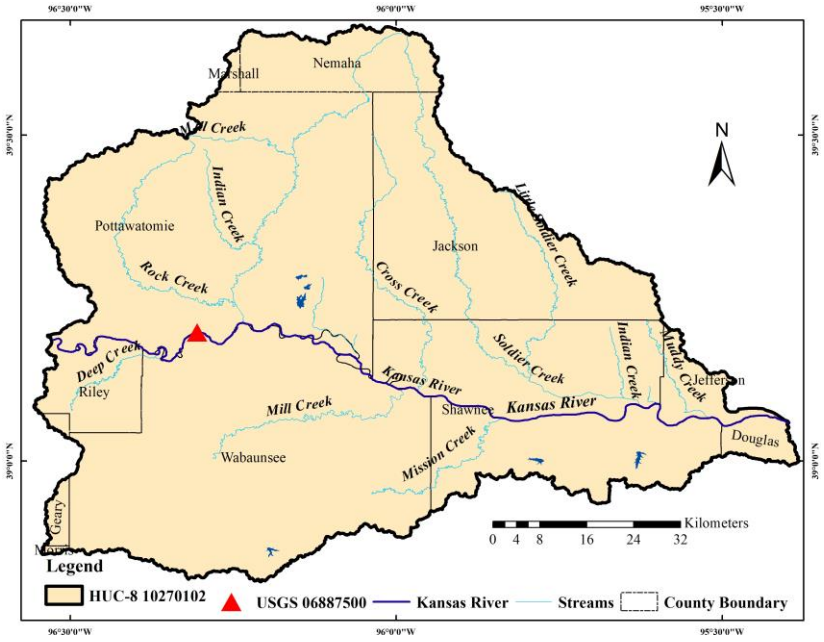


Figure 3.3 Location of the Kansas River at Wamego (HUC-8 10270102). Red triangle shows a continuous water quality monitoring station (USGS 06887500).

3.2 Methodology

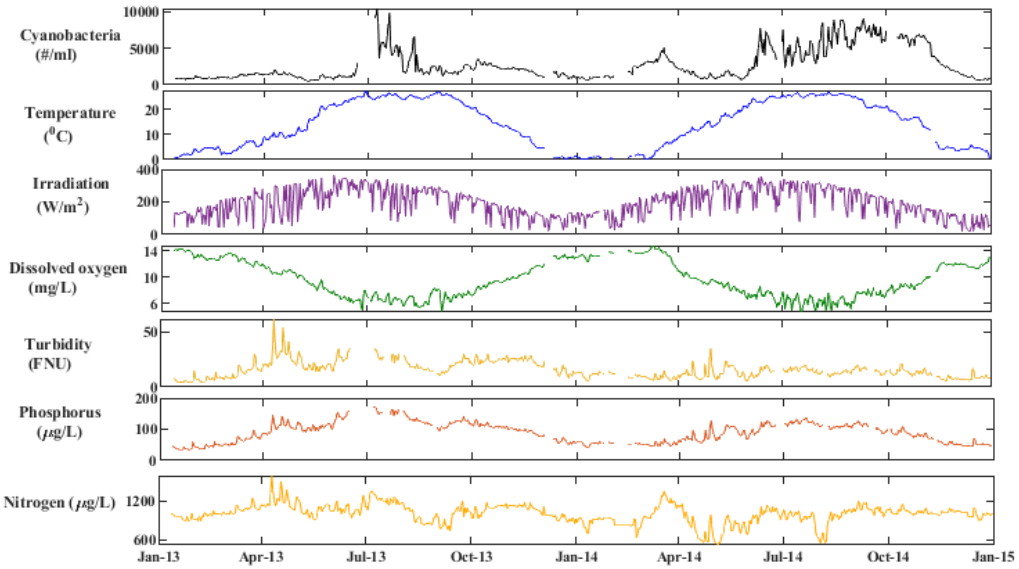
3.2.1 Data Collection

Two years (2013-2014) of sub-daily data for eight variables, including weather, environmental factors, and cyanobacteria concentrations, were collected at the USGS station 07144790 in Cheney Reservoir (Figure 3.4) (https://nrtwq.usgs.gov/). Phosphorus and nitrogen

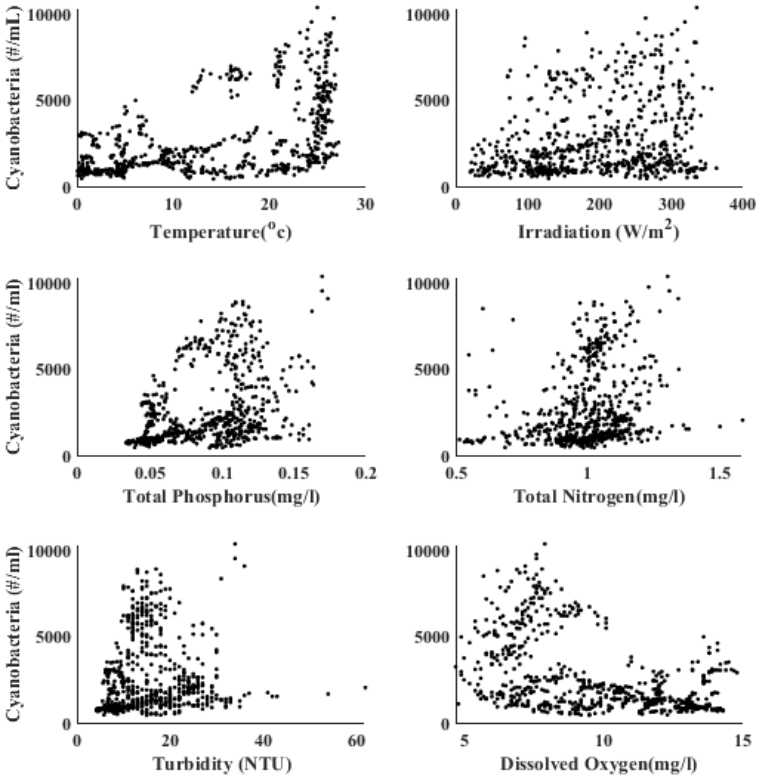
concentrations were based on the regression model developed by the USGS and presented in Appendix. Global Horizontal Irradiation (GHI) data for both of the study sites were collected from the National Solar Radiation Database. Daily data of temperature, irradiation, turbidity, dissolved oxygen, and cyanobacteria concentrations were collected for the Kansas River (Wamego) at station USGS 06887500 for 2012-2014 (Figure 3.5) and listed for 15-day intervals in Appendix A.

Notation	Definition	Continuous monitoring (C) or Modeled (M) data	Mean	Standard Deviation	Observed Range/ Comments
Direct Variables					
C	Cell concentration of cyanobacteria (cells/ml)	C	2789	2172.3	509 - 10400
T	Temperature(℃)	C	13.85	9.33	0.1 - 27.7
I	Global irradiance (GHI) (W/m^2)	C	193.98	87.01	20 - 364
P_{tot}	Total phosphorus concentration ($\mu g\ L^{-1}$)	M	90.5	33.3	34 - 174
N_{tot}	Total nitrogen concentration ($\mu g\ L^{-1}$)	M	995	147.4	515 - 1587
q_{in}	Inflow volume of lake ($m^3 day^{-1}$)	C	277417	743098	0 - 10520294
V	volume of lake (acre-ft)	C	151328	27069	101500 - 218400
Indirect Variables					
	Lake depth (m)	C	8.17	0.972	6.3 - 10.24
	Turbidity (FNU)	C	16.05	7.69	4.2 - 62
	Dissolved Oxygen (mg/L)	C	9.75	2.73	4.7 - 14.8

(a) Summary statistics in Cheney Reservoir



(b) Time-series data in Cheney Reservoir

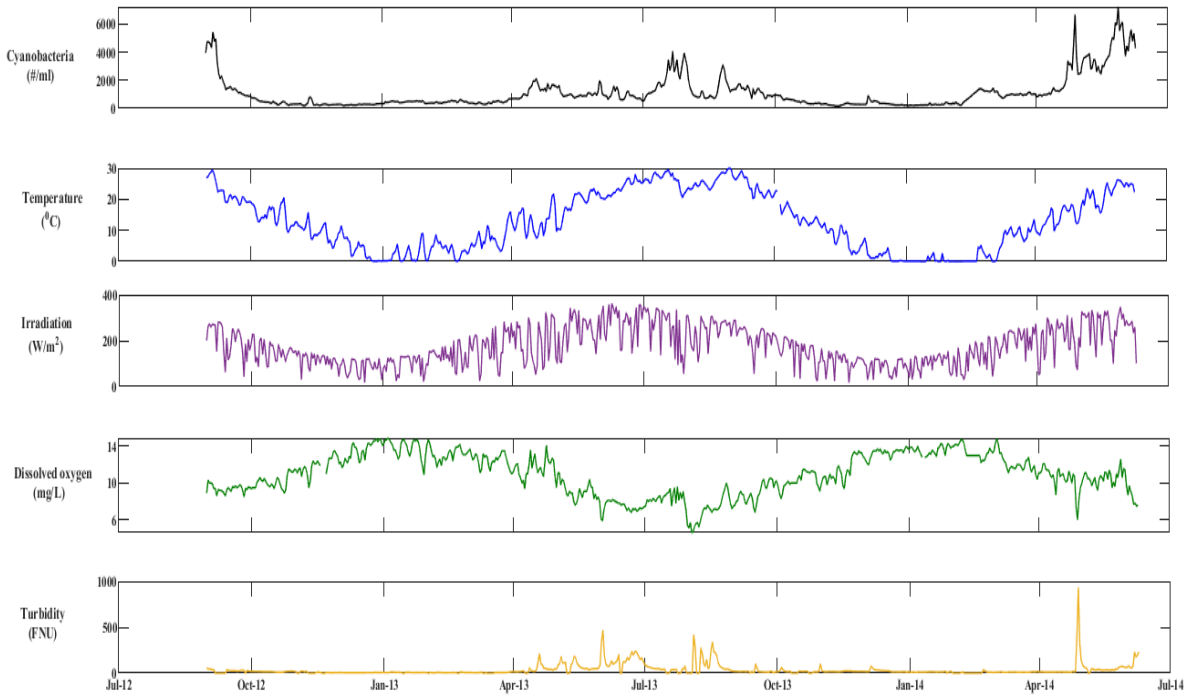


(c) Daily cyanobacteria concentration vs. six environmental factors in Cheney Reservoir

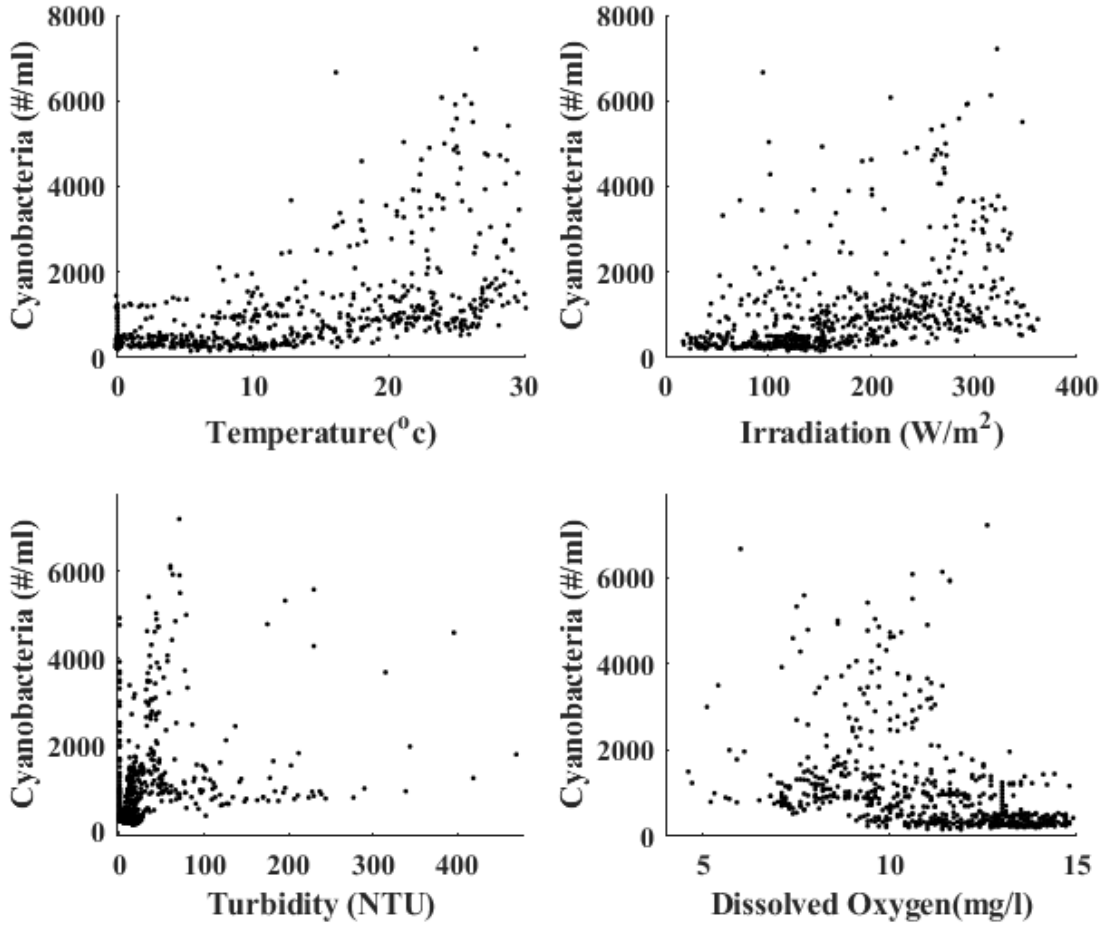
Figure 3.4 Cheney Reservoir data summary during 2013-2014.

Notation	Definition	Continuous monitoring (C) or Modeled (M)	Mean	Standard Deviation	Observed Range/ Comments
Direct Variables					
C	Cell concentration of cyanobacteria (cells/ml)	C	1090	1156	160 - 7212
T	Temperature(℃)	C	12.73	9.28	-0.1 - 30.1
I	Global irradiance (GHI) (W/m^2)	C	175.5	84.53	17.37 - 361.91
q_{in}	River flow ($m^3 day^{-1}$)	C	4269274	6791694	1196000 - 56760553
Indirect Variables					
	Lake depth (m)	C	1.26	0.35	0.95 - 3.55
	Turbidity (FNU)	C	35.53	67.02	0 - 470
	Dissolved Oxygen (mg/L)	C	10.91	2.41	4.6 - 14.9

(a) Summary statistics in Kansas River at Wamego



(b) Time-series data in Kansas River at Wamego



(c) Cyanobacteria concentration vs. environmental factors in Kansas River at Wamego
Figure 3.5 Kansas River at Wamego data summary during 2012-2014.

Summary statistics of the collected variables for both sites are included in Figure 3.4 (a) and Figure 3.5 (a). Analysis of the collected data during 2013-2014 in the Cheney Reservoir showed an average cyanobacteria concentration of 2789 cells/ml, a maximum of 10,400 cells/ml and a minimum of 509 cells/ml. The range in concentration shows that cyanobacteria was always over the '*vigilance level* (>200 cells/ml)' and frequently exceeded the '*alert level 1* (>10,000)' for drinking water concern. Kansas River at Wamego had an average cyanobacteria concentration of 1090 cells/ml during the study period. The dissolved oxygen level was on

average at 9.75-10.91 mg/l with a minimum of 4.6-4.7 mg/l at both sites. The observed minimum concentration for dissolved oxygen can be critical for aquatic life, which requires for survival ≥ 5 mg/l during the warm season (KDHE, 2011, pp. 1-3).

3.2.2 Correlation Analysis

The correlation analysis was conducted to evaluate the interactions between cyanobacteria concentration and environmental variables like temperature, irradiation, dissolved oxygen, turbidity, phosphorus, and nitrogen (Helsel and Hirsch, 2002, pp. 210-218). A correlation matrix with the Pearson's *r* (Pearson Product Moment Coefficient) was developed to find multicollinearity between the variables using equation (3.1):

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right) \tag{3.1}$$

Where, x_i, y_i= *n* data pairs for variables x and y

$\bar{x}, \bar{y}$= mean value for x and y

s_x, s_y=standard deviation for x and y

Pearson's r is a dimensionless coefficient that varies from -1 to 1. r=0 indicates no correlation between two variables, while *r* close to ± 1 indicates a strong relationship. Positive r suggests both variables tend to increase in the same direction, whereas the negative r indicates opposite trends in correlated variables.

3.2.3 Correlation Matrix

The data were analyzed using the Pearson correlation coefficient for the Cheney Reservoir (Table 3-2) and Kansas River (Table 3-3). Cyanobacteria concentration was found to be positively correlated with water temperature and irradiation for both of the aquatic systems. At both sites,

cyanobacteria was found negatively correlated with dissolved oxygen. This appears reasonable as cyanobacteria normally requires oxygen to be absorbed from aquatic bodies for photosynthesis. Thus, cyanobacteria's proliferation results in depletion of oxygen, which is harmful to the aquatic ecosystem. Cyanobacteria positively correlated with phosphorus concentration, nitrate plus nitrite concentration, water elevation, and reservoir storage in Cheney Reservoir. For Kansas River, cyanobacteria was found positively correlated with turbidity and chlorophyll-a. Temperature was found significantly correlated with irradiation, phosphorus, and turbidity in both water bodies.

Table 3-2 Pearson correlation coefficient of cyanobacteria concentration with sixteen environmental variables for Cheney Reservoir. Bold-underlined indicates a statistically significant association for cyanobacteria concentration with a p-value of less than 0.05.

<-0.40 | <|±0.40| | >0.40

Cheney Reservoir

	Cyanobacteria	Water temperature	Irradiation (GHI)	Total Phosphorus (TP)	Total Nitrogen (TN)	TN/TP	Turbidity	Elevation	Discharge (Inflow)	Reservoir Volume	Dissolved Oxygen	Chlorophyll-a	pH	Irradiation (DNI)	Nitrate plus Nitrite	Nitrate plus Nitrite/TP
Cyanobacteria	1.00	**0.54**	**0.42**	**0.53**	0.20	**-0.47**	0.24	**0.40**	-0.14	**0.39**	**-0.64**	-0.10	0.12	**0.39**	**0.57**	**0.59**
Water temperature	**0.54**	1.00	0.77	0.84	0.18	-0.81	0.54	0.16	-0.03	0.15	-0.89	-0.12	-0.10	0.54	0.88	0.87
Irradiation (GHI)	**0.42**	0.77	1.00	0.71	-0.04	-0.74	0.44	0.16	0.12	0.16	-0.73	0.06	-0.29	0.79	0.72	0.68
Phosphorus	**0.53**	0.84	0.71	1.00	0.27	-0.94	0.85	0.19	0.13	0.18	-0.89	-0.30	-0.25	0.48	0.98	0.90
Nitrogen	0.20	0.18	-0.04	0.27	1.00	0.07	0.43	-0.40	-0.10	-0.41	-0.03	0.24	0.77	-0.08	0.17	0.02
N/P	**-0.47**	-0.81	-0.74	-0.94	0.07	1.00	-0.73	-0.34	-0.17	-0.34	0.91	0.39	0.53	-0.53	-0.96	-0.93
Turbidity	0.24	0.54	0.44	0.85	0.43	-0.73	1.00	0.01	0.15	-0.01	-0.53	-0.20	-0.21	0.31	0.76	0.61
Elevation	**0.40**	0.16	0.16	0.19	-0.40	-0.34	0.01	1.00	0.25	1.00	-0.32	-0.27	-0.38	0.34	0.26	0.35
Discharge (In flow)	-0.14	-0.03	0.12	0.13	-0.10	-0.17	0.15	0.25	1.00	0.25	-0.08	-0.12	-0.27	0.03	0.10	0.05
Reservoir Volume	**0.39**	0.15	0.16	0.18	-0.41	-0.34	-0.01	1.00	0.25	1.00	-0.32	-0.27	-0.38	0.34	0.26	0.34
Dissolved Oxygen	**-0.64**	-0.89	-0.73	-0.89	-0.03	0.91	-0.53	-0.32	-0.08	-0.32	1.00	0.37	0.30	-0.50	-0.95	-0.98
Chlorophyll-a	-0.10	-0.12	0.06	-0.30	0.24	0.39	-0.20	-0.27	-0.12	-0.27	0.37	1.00	0.41	-0.01	-0.35	-0.41
pH	0.12	-0.10	-0.29	-0.25	0.77	0.53	-0.21	-0.38	-0.27	-0.38	0.30	0.41	1.00	-0.26	-0.29	-0.34
Irradiation (DNI)	**0.39**	0.54	0.79	0.48	-0.08	-0.53	0.31	0.34	0.03	0.34	-0.50	-0.01	-0.26	1.00	0.50	0.50
Nitrate plus Nitrite	**0.57**	0.88	0.72	0.98	0.17	-0.96	0.76	0.26	0.10	0.26	-0.95	-0.35	-0.29	0.50	1.00	0.97
Nitrate plus Nitrite/P	**0.59**	0.87	0.68	0.90	0.02	-0.93	0.61	0.35	0.05	0.34	-0.98	-0.41	-0.34	0.50	0.97	1.00

Table 3-3 Pearson correlation coefficient of cyanobacteria concentration with eight environmental variables for the Kansas River (Wamego). Bold-underlined indicates a statistically significant association for cyanobacteria concentration with a p-value of less than 0.05.

<-0.40 <|±0.40| >0.40

Kansas River (Wamego)

	Cyanobacteria	Water temperature	Irradiation (GHI)	Turbidity	Elevation	Discharge	Dissolved Oxygen	Chlorophyll-a	pH
Cyanobacteria	1.00	**0.54**	**0.81**	**0.53**	0.16	0.27	**-0.44**	**0.69**	-0.01
Water temperature	**0.54**	1.00	0.75	0.52	0.07	0.29	-0.72	0.61	0.29
Irradiation (GHI)	**0.81**	0.75	1.00	0.69	0.32	0.45	-0.76	0.62	0.17
Turbidity	**0.53**	0.52	0.69	1.00	0.67	0.77	-0.75	0.43	0.15
Elevation	0.16	0.07	0.32	0.67	1.00	0.96	-0.59	-0.26	-0.17
Discharge (In flow)	0.27	0.29	0.45	0.77	0.96	1.00	-0.69	-0.10	-0.10
Dissolved Oxygen	**-0.44**	-0.72	-0.76	-0.75	-0.59	-0.69	1.00	-0.27	-0.15
Chlorophyll-a	**0.69**	0.61	0.62	0.43	-0.26	-0.10	-0.27	1.00	0.38
pH	-0.01	0.29	0.17	0.15	-0.17	-0.10	-0.15	0.38	1.00

3.2.4 Model Statement

Cyanobacteria is a sub-group of phytoplankton that performs photosynthesis using sunlight by its cell pigments (Chorus and Bartram, 1999, para 2.2.1). Cyanobacteria functions similarly to the algal community in an eco-aquatic system (Graham et al., 2017, p. 3). This suggests that the phytoplankton model can be used as a baseline for mechanistic modeling of the cyanobacteria population. Fundamentals of the phytoplankton dynamics and model development for cyanobacteria growth are described in this section.

3.2.4.1 Phytoplankton Model

Phytoplankton is a free-floating microorganism in the aquatic system. Phytoplankton abundance is controlled by four major processes: (i) growth, (ii) non-predatory loss, (iii) grazing

loss, and (iv) washout loss by the flow. The rates related to growth (GR), non-predatory loss (NL), grazing loss (GL), and washout (WO) control the net growth rate (NGR) of the phytoplankton in a waterbody. The waterbody can be considered as a continuously stirred tank reactor (CSTR) that assumes uniform spatial variability of the microbes (Bowie et al., 1985, p. 13; Chapra, 1997, p. 614; Malve et al., 2007, p. 968). Various types of phytoplankton exist in the aquatic systems. Since these types respond differently to the same environment, three major species like diatoms, green algae, and cyanobacteria are commonly considered distinct phytoplankton groups in phytoplankton modeling (Bowie et al., 1985, p. 280).

A general mass balance equation for *n* phytoplankton groups (A_i) that represent dynamics of phytoplankton growth can be stated as a first-order kinetic reaction equation (*n* phytoplankton groups are indicated as *i*= 1, 2, 3 ...*n*):

$$\frac{dA_i}{dt} = (GR_i - NL_i - GL_i - WO_i)A_i \qquad (3.2)$$

Different phytoplankton groups compete against each other exclusively for nutrients availability. Each of these phytoplankton groups uptakes phosphorus, nitrogen, and other macro-nutrients from the aquatic system. However, their internal nutrient content (α_i, β_i) is not available to any other group that affects nutrient limitations for the growth rate. So, the amount of nutrients available for phytoplankton is lower than the total amount of nutrients in the reservoir. Interaction among different groups for nutrient availability and effects of environmental factors ($\varphi_{Temp_i}, \varphi_{Nutrient_i}, \varphi_{Light_i}$ in Table 2-2) on the growth rate *GR* of each of the species can be modeled as equation (3.2a).

$$GR_i = \mu_{m_i}\varphi_{Temp_i}\varphi_{Nutrient_i}\varphi_{Light_i} \qquad (3.2a)$$

Where, $\varphi_{Nutrient_i} = \frac{P}{K_{P_i}+P}\frac{N}{K_{N_i}+N}$

$$P = P_{tot} - \sum_{i=1}^{n} \alpha_i A_i \; and \; N = N_{tot} - \sum_{i=1}^{n} \beta_i A_i$$

Growth rate GR_i of a phytoplankton group can be different for each group in the same environment due to their variability in response to temperature, light, zooplankton grazing, and internal nutrient requirements. Their internal nutrient content controls the growth rate of each phytoplankton group in the system. During a bloom event one group may dominant the other groups and in that case ignoring minor groups can be considered. In the current study, continuous monitoring data for all the phytoplankton groups were not available, but during periods of HABs cyanobacteria composed a dominant group. Thus, we assumed the effect of minor algal groups negligible. In the following section, cyanobacteria modeling approach based on the phytoplankton model (3.2) is presented.

3.2.4.2 Cyanobacteria Model

Cyanobacteria modeling approach used in this study is based on a conceptual phytoplankton dynamics model described in section 3.2.4.1. Cyanobacteria is assumed a dominant specie in the studied waterbody, and its concentration is denoted by *C*. Zooplankton grazing effect is not well understood for controlling cyanobacteria (Pearl and Otten, 2013). Grazing effect can be important during bloom initiation period and under oligotrophic conditions. However, during bloom periods and in eutrophic lake conditions grazing pressure usually is minimal. Since this study deals with eutrophic water bodies and no zooplankton data were available during the study period, zooplankton grazing effect (GL) is considered insignificant (based on Chow-Fraser (1994, p. 2052), Chapra et al. (2017, p. 8936)) and the grazing loss term *GL* in equation (3.2) is dropped. Re-writing the equation (3.2) for cyanobacteria concentration leads to

$$\frac{dC}{dt} = (GR - NL - WO)C \tag{3.3}$$

We define the net growth rate (NGR) of cyanobacteria as a sum of its growth and two loss terms in equation (3.4):

$$NGR = GR - NL - WO \tag{3.4}$$

The terms in equation (3.4) are the functions of the variables listed in Table 3-4. In equation (3.3), using concentration, *C(t)*, and driving factors (*T, I, N, P,* q_{out}, *V*) at time *t*, concentration *C(t+Δt)* can be found at the time *t+Δt* by integration. The schematic diagram of solving the equation (3.3) for a single time step *Δt* is presented in Figure 3.6.

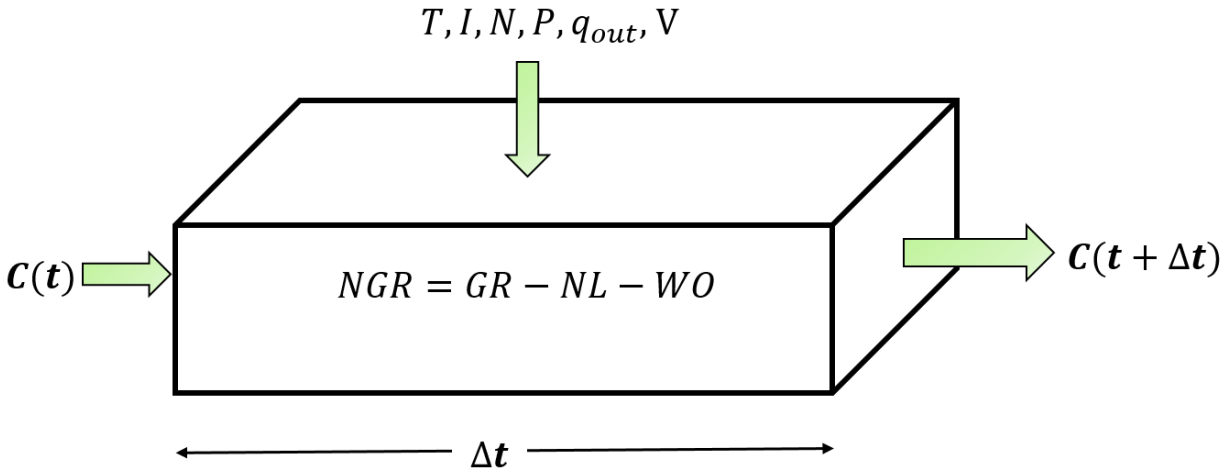


Figure 3.6 Conceptual cyanobacteria model

Growth Rate:

Phytoplankton growth rate (GR) is a function of temperature, light, and nutrients (Bowie et al., 1985, p. 287; Chapra, 1997, p. 604) and can be expressed as:

$$GR = \mu_m \varphi_{Temp} \varphi_{Nutrient} \varphi_{Light} \tag{3.5}$$

Functional forms of the growth factors for temperature (φ_{Temp}), nutrients ($\varphi_{Nutrient}$) and light (φ_{Light}) in equation (3.5) were discussed in Chapter 2 (Table 2-2), whereas their parameters and variables are presented in Table 3-4.

Table 3-4 Notations and definition of parameters and variables

	Definition	Unit
Parameters		
μ_m	maximum growth rate at 20℃	day^{-1}
σ_m	maximum non-predatory loss rate at 20℃	day^{-1}
θ_μ	temperature coefficients for growth rate	-
θ_σ	temperature coefficients for non-predatory loss rate	-
K_I	global irradiance half-saturation coefficient	$W\ m^{-2}$
K_P	phosphorus half-saturation coefficient	$\mu g\ L^{-1}$
K_N	nitrogen half-saturation coefficient	$\mu g\ L^{-1}$
α	relative phosphorus content of cyanobacteria	-
β	relative nitrogen content of cyanobacteria	-
φ_{Temp}	temperature factor on the maximum growth rate at a reference temperature	-
$\varphi_{Nutrient}$	attenuation factors for nutrients	-
φ_{Light}	attenuation factors for light	-
Variables		
A	cell concentration of phytoplankton	$cells\ L^{-1}$
C	cell concentration of cyanobacteria	$cells\ L^{-1}$
P	available phosphorus for cyanobacteria	$\mu g\ L^{-1}$
P_{tot}	total phosphorus concentration	$\mu g\ L^{-1}$
N	available nitrogen for cyanobacteria	$\mu g\ L^{-1}$
N_{tot}	total nitrogen concentration	$\mu g\ L^{-1}$
T, T_{ref}	temperature, the reference temperature (20℃)	℃
q_{out}	outflow volume	$m^3 day^{-1}$
I	global irradiance	$W\ m^{-2}$
V	volume of lake	m^3

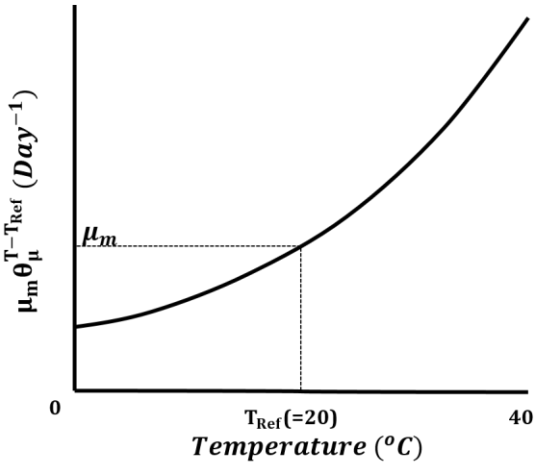


Figure 3.7 Effect of temperature on the growth rate: theta model (adapted from Bowie et al., 1985, p. 294)

The choice of functional forms of $\varphi_{Temp}, \varphi_{Nutrient}, \varphi_{Light}$ for cyanobacteria from Table 2-2 is based on the scope of a study. In this study, we use the theta model (type I of φ_{Temp} in Table 2-2) based on Arrhenius or van't Hoff equation to consider the effect of temperature on the growth rate (Bowie et al., 1985, p. 255). Theta model simulates cyanobacteria growth rate at a specific temperature (T) based on the maximum growth rate (μ_m) of cyanobacteria in nutrients and light-saturated environment at a reference temperature, T_{ref} , often 20 ^oC (Figure 3.7):

$$\varphi_{Temp} = \theta_{\mu}^{T-T_{Ref}} \qquad (3.5a)$$

This model can also be used when considering single phytoplankton species below the optimum temperature of growth. Chapra et al. (2017) used 29.7 ^oC as cyanobacteria optimum temperature. In our study, the water temperature was always below 29 ^oC for the Cheney Reservoir and 31 ^0C for the Kansas River, which justifies using type I form for φ_{Temp}.

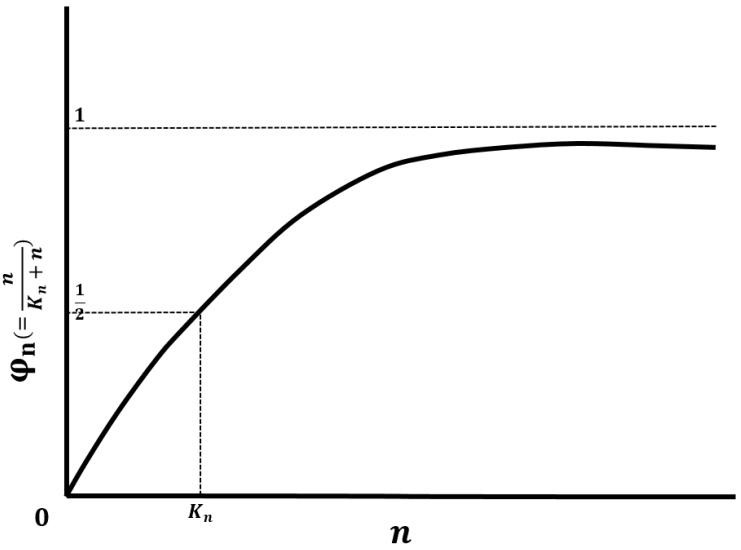


Figure 3.8 Effect of nutrients and solar irradiation on the growth rate using the Monod model (adapted from Chapra, 1997, p. 607)

Phosphorus (P), nitrogen (N), carbon, and silicon are the limiting nutrients considered in the phytoplankton model, but their selection depends on a phytoplankton group (Bowie et al., 1985, pp. 287-325). For cyanobacteria, silicon and carbon are not considered limiting. Carbon is

frequently available in the environment, whereas silicon can be regarded as limiting for diatom groups. Available nutrients for cyanobacteria (P, N) and light (I) limitation effects are simulated using the Monod kinetics (Monod, 1949). This assumes first-order kinetics for concentrations up to half-saturation coefficients (k_n), while zero-order kinetics (Figure 3.8) applies to higher levels of nutrient concentrations or light intensity:

$$\varphi_n = \frac{n}{K_n + n} \tag{3.5b}$$

where *n* represents either available phosphorus (P), or available nitrogen (N), or light intensity (I).

After substituting (3.5a) and (3.5b) into (3.5), the growth rate function in equation (3.5) is transformed to

$$GR = \mu_m \theta_\mu^{T-T_{Ref}} \frac{P}{K_P + P} \frac{N}{K_N + N} \frac{I}{K_I + I} \tag{3.6}$$

Where, $P = P_{tot} - \alpha C$ and $N = N_{tot} - \beta C$.

Natural Loss:

Natural loss (NL) is defined as a non-predatory loss of cyanobacteria. Non-predatory loss considers the effects of respiration-excretion and cell mortality due to viral lysis or stressed natural environment except zooplankton grazing. It is often modeled as a function of temperature using the theta model as in equation (3.7) (Bowie et al., 1985, p. 351; Chapra, 1997, p. 614):

$$NL = \sigma_m \theta_\sigma^{T-T_{Ref}} \tag{3.7}$$

Washout:

Cyanobacteria can be carried away from the reservoir with streamflow, and it is called a washout. In-lake cyanobacteria washout in a CSTR reservoir can be modeled with stream outflow (q_{out}) and storage (V) as

$$WO = \frac{q_{out}}{V} \tag{3.8}$$

To summarize, our assumptions for cyanobacteria model development are:

1. Cyanobacteria is a dominant phytoplankton specie in a waterbody
2. The reservoir is modeled as a continuously stirred tank reactor (CSTR)
3. Theta model is used for water temperature (T) effect on growth rate (GR) and non-predatory loss rate (NL)
4. Michaelis-Menten kinetics is used to describe the limiting effects of some variables (nutrition/light) on the growth rate (GR)
5. Predatory loss due to zooplankton grazing is assumed minimal.

3.2.4.3 Formulation of Cyanobacteria Models

Continuous monitoring and collection of cyanobacteria concentrations and environmental variables from Equation (3.6) can be time-consuming and costly. Alternatively, models with fewer variables are valuable in a data-scarce context. In this study, we propose four alternative models shown in Table 3-5 for growth term (GR) in equation (3.6) based on combinations of four GR factors: Model 1 (T-based), Model 2 (T, I- based), Model 3 (T, I, P- based), and Model 4 (T, I, P, N- based). Each consecutive model increases the number of driving factors by one. Model 1 considers only temperature, while Model 4 is the most comprehensive with all four factors included.

Table 3-5 Types of proposed alternative models and driving factors affecting GR

Model type	GR factors	GR function
Model 1	T	$\mu_m \theta_\mu^{T-T_{Ref}}$
Model 2	T, I	$\mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I}$
Model 3	T, I, P	$\mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I} \frac{P}{K_P + P}$
Model 4	T, I, P, N	$\mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I} \frac{P}{K_P + P} \frac{N}{K_N + N}$

3.2.4.4 Parameter Estimation

Equation (3.3) can be solved by integrating both sides of the equation and assuming the NGR constant during a particular simulation period (Δt) as

$$C(t+\Delta t) = C(t)e^{NGR\left(t+\frac{\Delta t}{2}\right)\Delta t} \tag{3.9}$$

Rearranging terms in equation (3.9), the observed net growth rate (NGR_{obs}) during a single time-step Δt can be derived using measured values of cyanobacteria concentration at two times as

$$NGR_{obs}\left(t+\frac{\Delta t}{2}\right) = \frac{1}{\Delta t} ln \frac{C(t+\Delta t)}{C(t)} \tag{3.10}$$

Combining equations (3.4) and (3.10) yields the *observed growth rate* (GR_{obs}) presented as:

$$GR_{obs}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right) = \frac{1}{\Delta t} ln \frac{C(t+\Delta t)}{C(t)} + \sigma_m \theta_\sigma^{T-T_{Ref}} + \frac{q_{out}}{V} \tag{3.11}$$

The *simulated growth rate* (GR_{model}) can be presented from the growth rate model in equation (3.6) for four sub-models as:

$$GR_{Model\,1}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right) = \mu_m \theta_\mu^{T-T_{Ref}} \tag{3.12a}$$

$$GR_{Model\,2}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right) = \mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I} \tag{3.12b}$$

$$GR_{Model\,3}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right) = \mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I}\frac{P}{K_P + P} \tag{3.12c}$$

$$GR_{Model\,4}\left(\mathrm{t}+\frac{\Delta \mathrm{t}}{2}\right) = \mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I}\frac{P}{K_P + P}\frac{N}{K_N + N} \tag{3.12d}$$

Where, $T = \frac{\sum_{i=t}^{t+\Delta t} T_i}{\Delta t}$

$$P = \frac{\sum_{i=t}^{t+\Delta t} P_i}{\Delta t}$$

$$N = \frac{\sum_{i=t}^{t+\Delta t} N_i}{\Delta t}$$

$$I = \frac{\sum_{i=t}^{t+\Delta t} I_i}{\Delta t}$$

Parameters in the simulated models (3.12a) – (3.12d) need to be adjusted to match the observed growth rate to parametrize the models. We used the least-square curve fitting method with the residual error in the model fitting procedure described as:

$$GR_{obs}\left(t + \frac{\Delta t}{2}\right) = GR_{model}\left(t + \frac{\Delta t}{2}\right) + \varepsilon_t \quad (3.13)$$

To optimize values of two (in (3.12a)) to five (in (3.12d)) parameters by using the equation (3.13), the Matlab curve-fitting tool '`lsqcurvefit`' was utilized. The tool minimized the sum of square errors (i.e., minimizing $\sum \varepsilon^2$) when the upper and lower bounds of each parameter were provided based on practical considerations and literature discussed in section 3.2.4.5. Once the parameters were optimized, the values were used in the model for forecasting cyanobacteria dynamics and presented in section 3.2.4.6.

3.2.4.5 Model Parameters

For model calibration, ranges for the values of nine model parameters were selected based on previous studies and presented in Table 3-6.

Table 3-6 Parameters range from past studies

Parameters	Literature Value	Unit	Source
μ_m	0.41 to 0.86 at 20℃	day^{-1}	Bowie et al. (1985)
	0.20 at 20℃		Pearl and Huisman (2009)
	0.9 at 29.7℃		Chapra et al. (2017)
σ_m	0.02 to 0.5	day^{-1}	Chapra (1997, p. 614)
	0 to 0.8		Bowie et al. (1985, p. 358)
θ_μ	1.01 to 1.2	-	Bowie et al. (1985)
	1.066		Eppley (1972)
θ_σ	1.08	-	Chapra (1997)
K_I	60	Ly/day	Chapra et al. (2017)
	0.002 to 0.004	$Kcal/m^2/sec$	Bowie et al. (1985)
	4-200	W/m^2	Malve et al. (2007)
K_P	2 to 60	$\mu g\ L^{-1}$	Bowie et al. (1985)

	1-5		Chapra et al. (2017)
K_N	15	μg L^{-1}	Chapra et al. (2017)
	1-20		Bowie et al.(1985)
	5-20		Chapra (1997)
	8-34		Malve et al. (2007)
α	0.013-0.035	-	Bowie et al.(1985, p. 334)
	0.004		Malve et al. (2007)
β	0.08-0.12	-	Bowie et al.(1985, p. 334)
	0.027		Malve et al. (2007)

3.2.4.6 Short-term forecast

For short-term forecasting, only one time-step calculation of Equation (3.9) was needed. At initial time t, the value of $C_{obs}(t)$ was taken from the observed dataset, and the values of the other physical variables (temperature, irradiation, and nutrient concentrations) were calculated as either average over the forecasting period or also taken at the initial time. The short-term forecasting formula is:

$$C_{model}(t+\Delta t) = C_{obs}(t)e^{NGR\left(t+\frac{\Delta t}{2}\right)\Delta t} \tag{3.14}$$

The forecasted value of $C_{model}(t+\Delta t)$ was compared with the observed value, and model efficiency was checked using available statistics, i.e., the Nash–Sutcliffe Efficiency (NSE) index and Root-Mean-Square Error (RMSE). NSE and RMSE for *D* days of daily data and a model simulation time-step of Δt (days) were calculated from the equations (3.15)-(3.16):

$$NSE = 1 - \frac{\sum_{t=0}^{n\Delta t}\left(C_{model}^{t+\Delta t} - C_{obs}^{t+\Delta t}\right)^2}{\sum_{t-0}^{n\Delta t}\left(C_{obs}^{t+\Delta t} - \overline{C_{obs}}\right)^2} \tag{3.15}$$

$$RMSE = \sqrt{\frac{\sum_{t=0}^{n\Delta t}\left(C_{model}^{t+\Delta t} - C_{obs}^{t+\Delta t}\right)^2}{\frac{D}{\Delta t} - 1}} \tag{3.16}$$

$$where, n = 1, 2, 3 \ldots \left(\frac{D}{\Delta t} - 1\right)$$

3.3 Results and Discussions

3.3.1 Model Simulation for Cheney Reservoir

3.3.1.1 Model Calibration and Backcasting

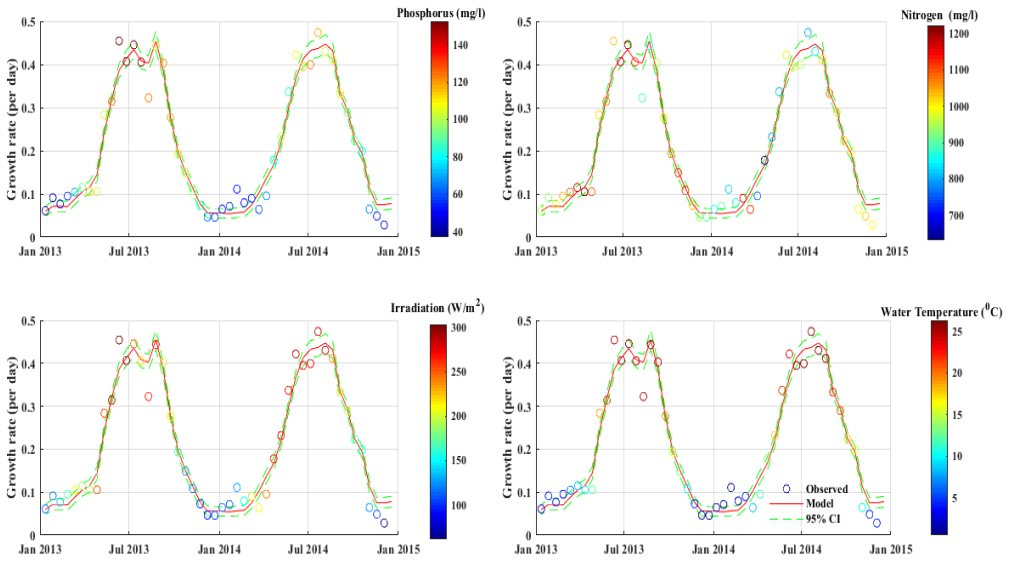
Four cyanobacteria models (Model 1, Model 2, Model 3, and Model 4) were calibrated based on the procedure presented in section 3.2.4.2 for the Cheney Reservoir. The calibrated values of model parameters and the associated statistics are provided in Table 3-7. The RMSE ranges from 1301 to 1364 cells/ml and NSE from 0.68 to 0.71 depending on the type of the model. A comparison of the observed growth rate time series for four models is shown in Figure 3.9.

Table 3-7. Model parameters with optimized values with 95% confidence interval (in brackets) for three simulation models in Cheney Reservoir ($\boldsymbol{\sigma = 0.27}$, $\boldsymbol{\theta_{\sigma} = 1.08}$)

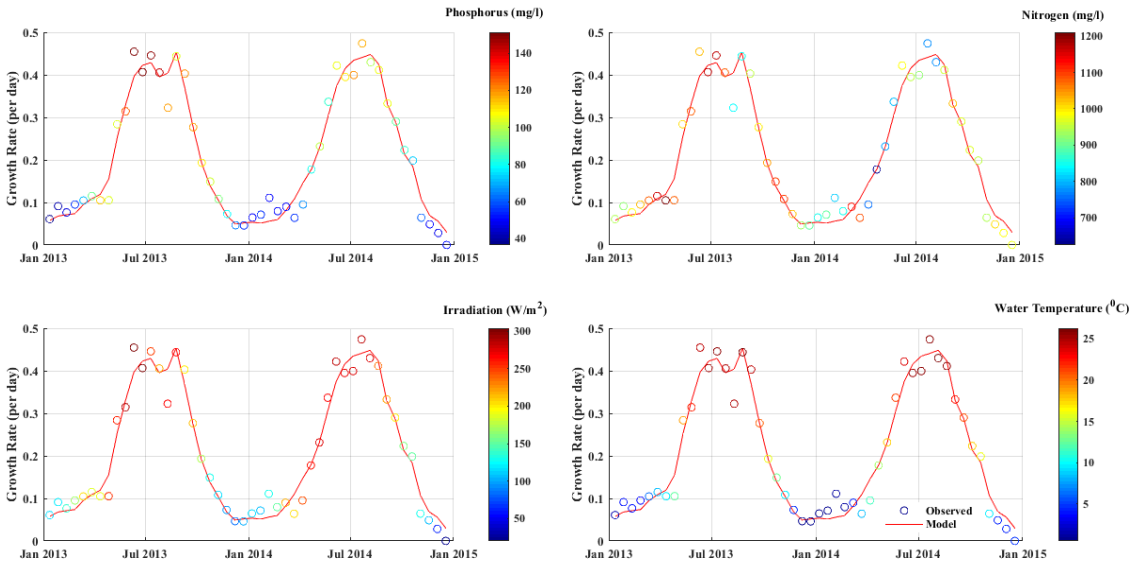
Parameters	μ_m (day^{-1})	θ_{μ}	K_I (W/m^2)	K_p ($\mu g\ L^{-1}$)	K_N ($\mu g\ L^{-1}$)	NSE (RMSE)
Cheney reservoir						
Model 1 (T)	0.2705 (0.26, 0.281)	1.086 (1.078, 1.094)	-	-	-	0.713 (1364)
Model 2 (T, I)	0.3322 (0.329, 0.335)	1.0794 (1.079, 1.0797)	49.73 (47.69, 51.78)	-	-	0.6825 (1341)
Model 3 (T, I, P) α=0.0953 (0.0948, 0.0957)	0.3316 (0.329, 0.333)	1.0786 (1.0783, 1.0789)	37.78 (35.98, 39.59)	2.37 (2.21, 2.525)	-	0.6457 (1240)
Model 4 (T, I, P, N) α=0.0953, β=0.12	0.3426 (0.339, 0.347)	1.0786 (1.0783,1.0789)	38.94 (37.06, 40.83)	2.25 (2.14, 2.35)	29.07 (21.42, 36.72)	0.6481 (1238)
Reference Value α=0.004-0.035, β=0.027-0.12	0.20-0.86	1.01-1.066	60-200	2-60	1-34	-

Fifteen-day cyanobacteria concentration time-series for a 15-day time-step simulations with calibrated model parameters are compared to observed data in Figure 3.10 and Figure 3.11. The simulated values are generated based on one-time-step simulations with initial cyanobacteria concentration at each 15-day time step taken from observed data and 15-day average values of the other environmental variables (temperature, irradiation, total phosphorus, and total nitrogen).

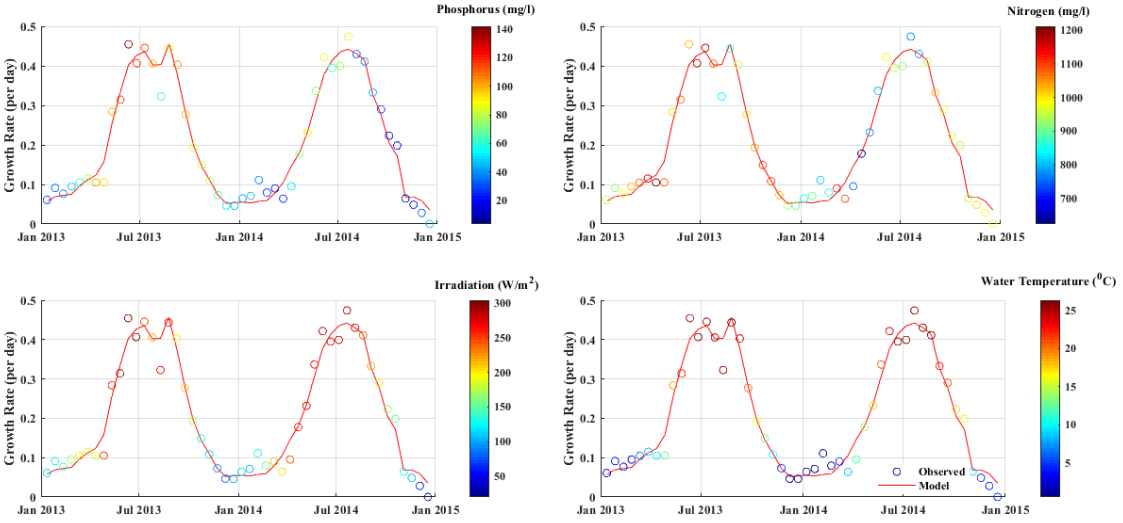
Optimized parameter values from Models 1 to 4 are in an acceptable range of reference values summarized in Table 3-7. Model 1, which provides a broader range of prediction intervals when simulating cyanobacteria concentration, shows a higher uncertainty level than the other models. At the same time, calibrated parameters ($\mu_m$and θ_μ) from Model 1 have a broader range of confidence intervals than in the other models. Optimized parameters calibrated from Model 3 has the least uncertainty in the range and, hence, lower uncertainty in simulating cyanobacteria concentration (Figure 3.10c). Values of NSE and RMSE from Table 3-7 show that Model 1 and Model 2 can be considered comparable in simulation performance with more complex Model 3 and Model 4. (T, I, P) - based Model 3 showed good long-term predictions of cyanobacteria concentration, comparable to the complete Model 4 (Figure 3.12a-b).



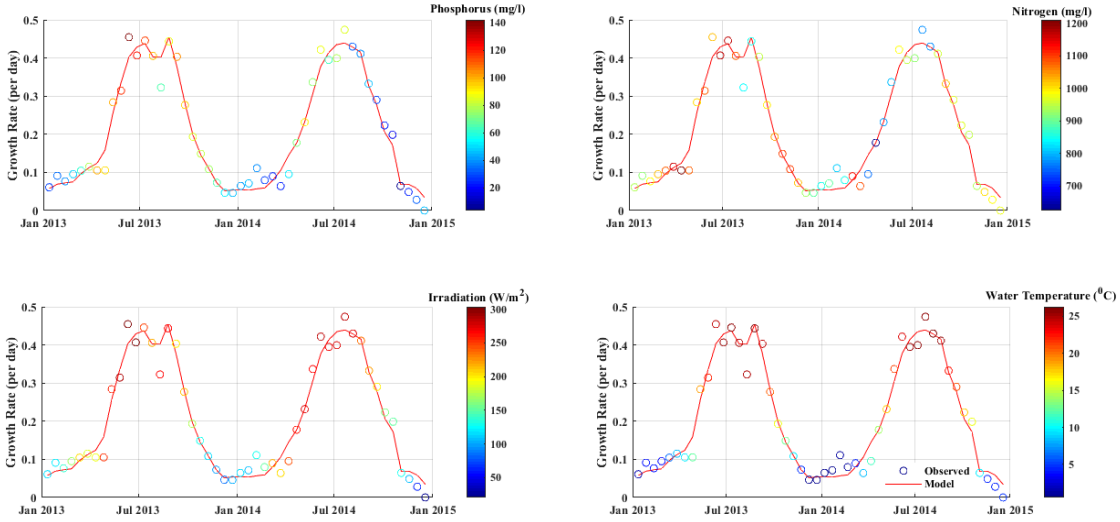
(a) Model 1



(b) Model 2

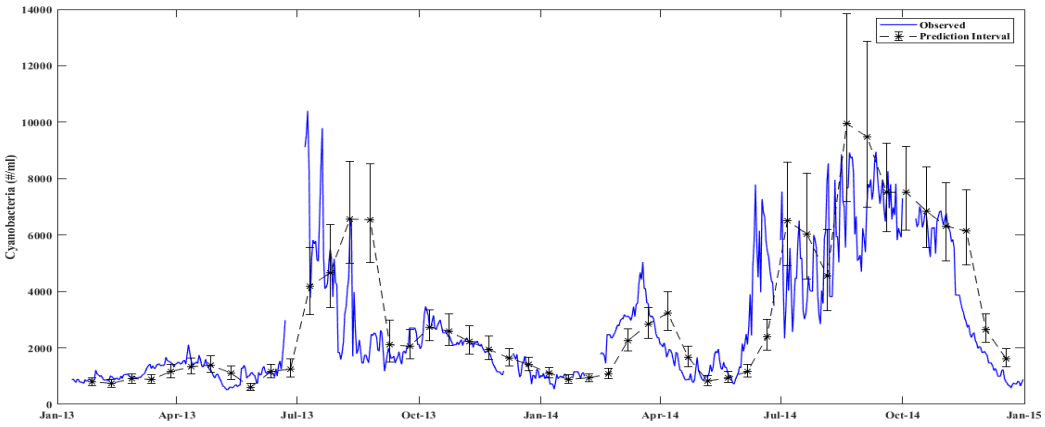


(c) Model 3

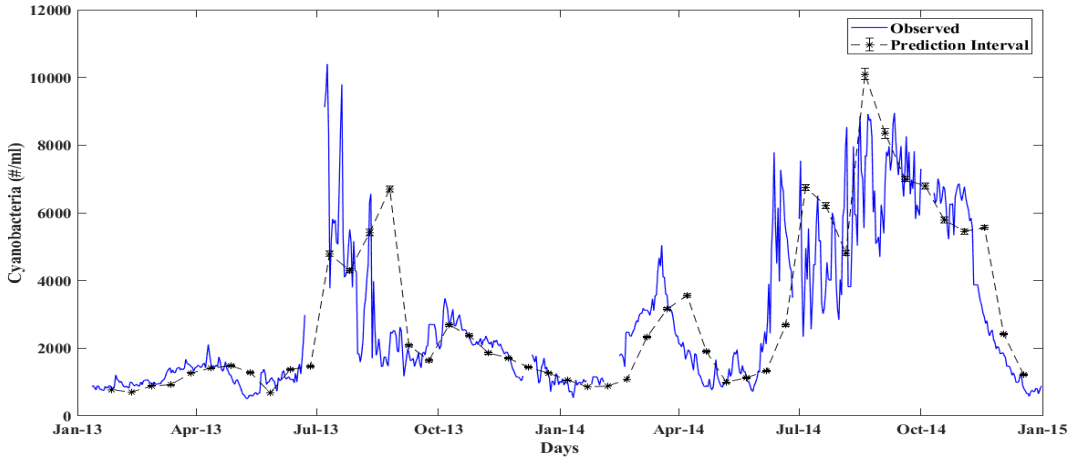


(d) Model 4

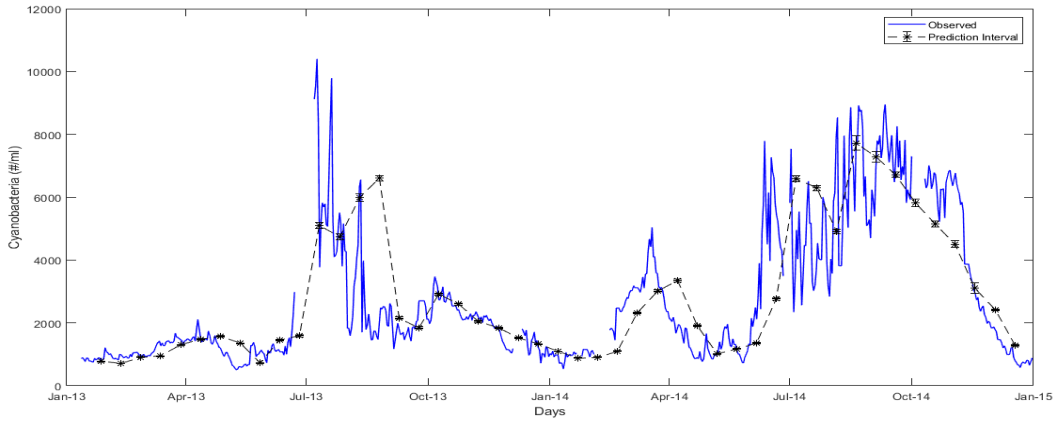
Figure 3.9 Graphical presentation of model calibration using 15 days averaged data sets in Cheney Reservoir based on (a) Model 1: R^2=0.9632; (b) Model 2: R^2=0.9659; (c) Model 3: R^2=0.9686 and (d) Model 4: R^2=0.9687. The color bar shows different environmental variables used in the model: available phosphorus (µg/l), available nitrogen (µg/l), global horizontal irradiation (W/m^2), and water temperature (^{o}C) respectively as indicated on top of the scale.



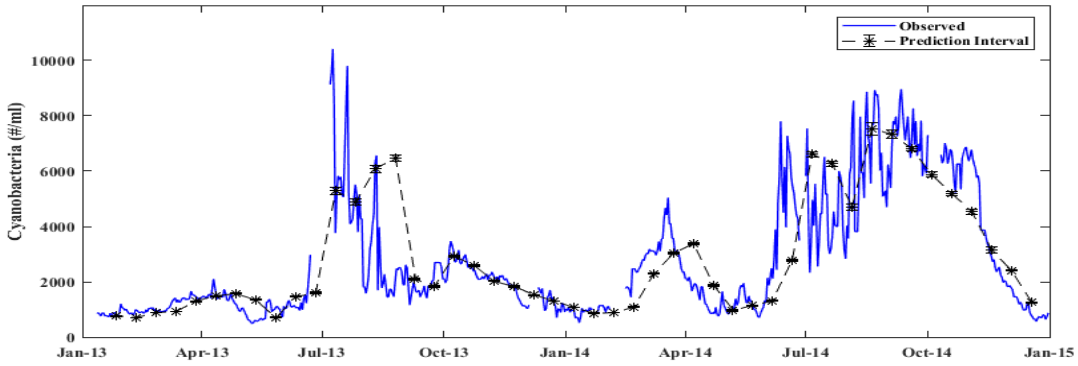
(a) Model 1



(b) Model 2



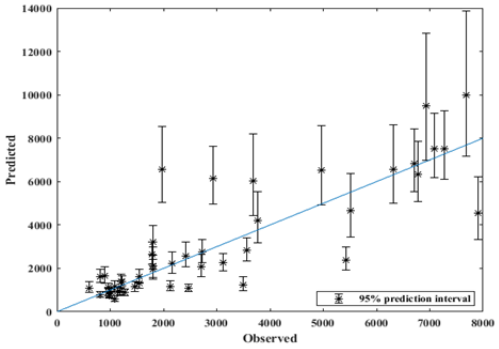
(c) Model 3



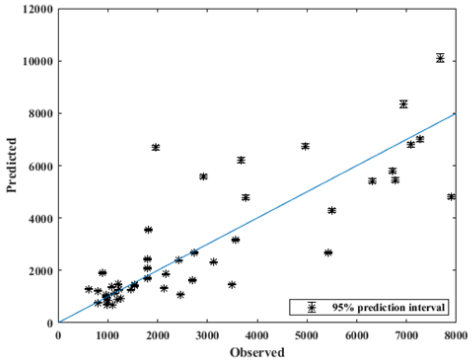
(d) Model 4

Figure 3.10 Model simulations for 15-day backcasting of cyanobacteria cell concentration in Cheney reservoir using (a) Model 1 (RMSE=1364, NSE=0.713), (b) Model 2 (RMSE=1341, NSE=0.6825), (c) Model 3 (RMSE=1240, NSE=0.6457), and (c) Model 4 (RMSE=1238, NSE=0.6481). The continuous blue line indicates observed daily

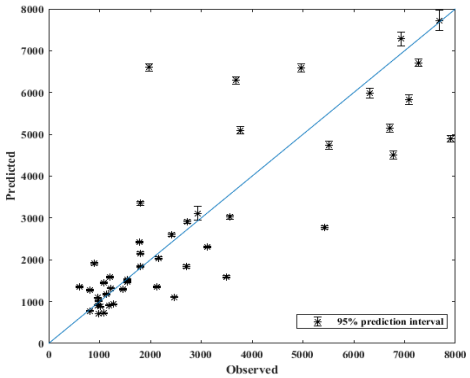
cyanobacteria concentrations in cells/ml, and the black dashed line indicates model outcomes using optimized parameters and environmental variables.



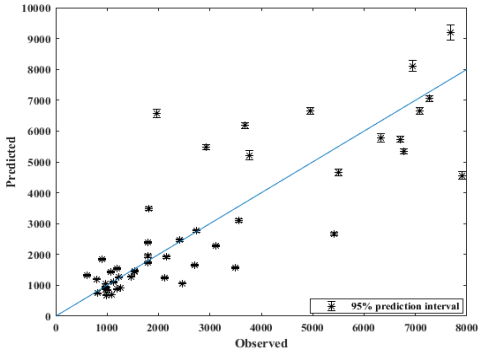
(a) Model 1



(b) Model 2

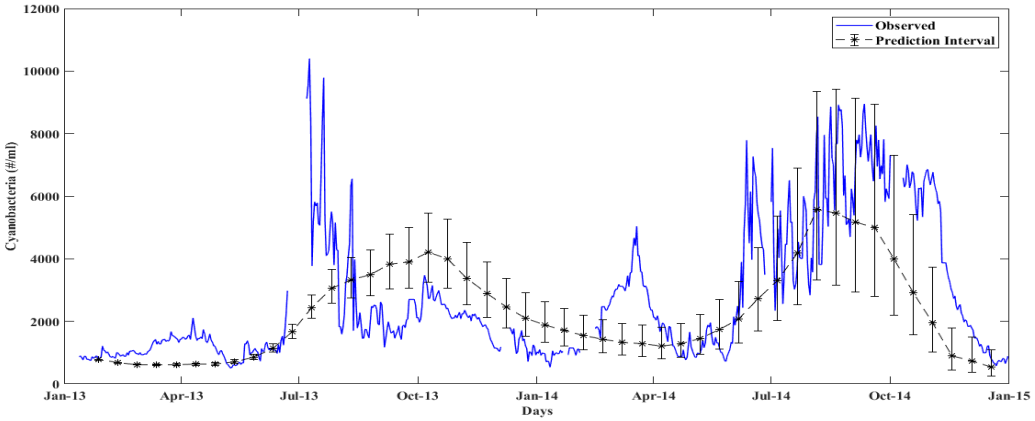


(c) Model 3

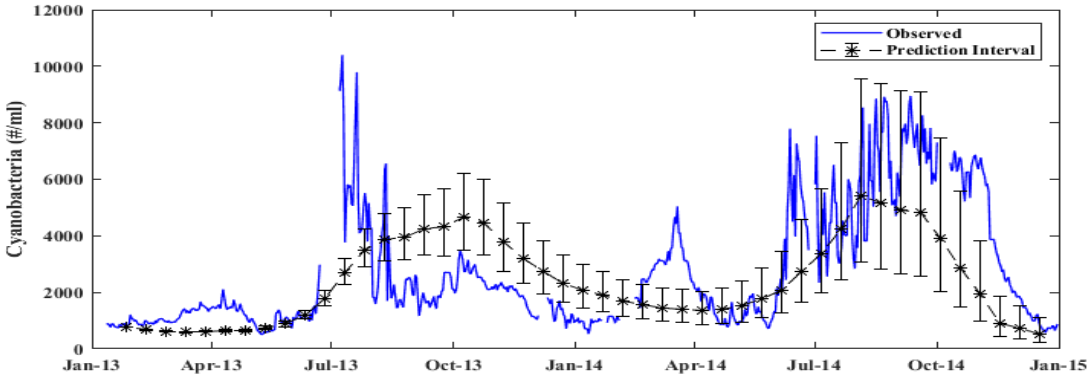


(d) Model 4

Figure 3.11 Simulated vs observed cyanobacteria concentration (#/ml) for CyanoHABs forecasting from (a) Model 1, (b) Model 2, (c) Model 3, and (d) Model 4 in Cheney reservoir. RMSE ranges from 1238 to 1364 in #/ml and NSE from 0.6457 to 0.713.



(a) Model 3



(b) Model 4

Figure 3.12 Model simulations for long-term backcasting of cyanobacteria cell concentration in the Cheney reservoir using (a) Model 3 and (b) Model 4. The continuous blue line indicates observed daily cyanobacteria concentrations in cells/ml, and the black dashed line indicates model outcomes using optimized parameters and environmental variables.

3.3.1.2 Time Step Effect on Model Performance

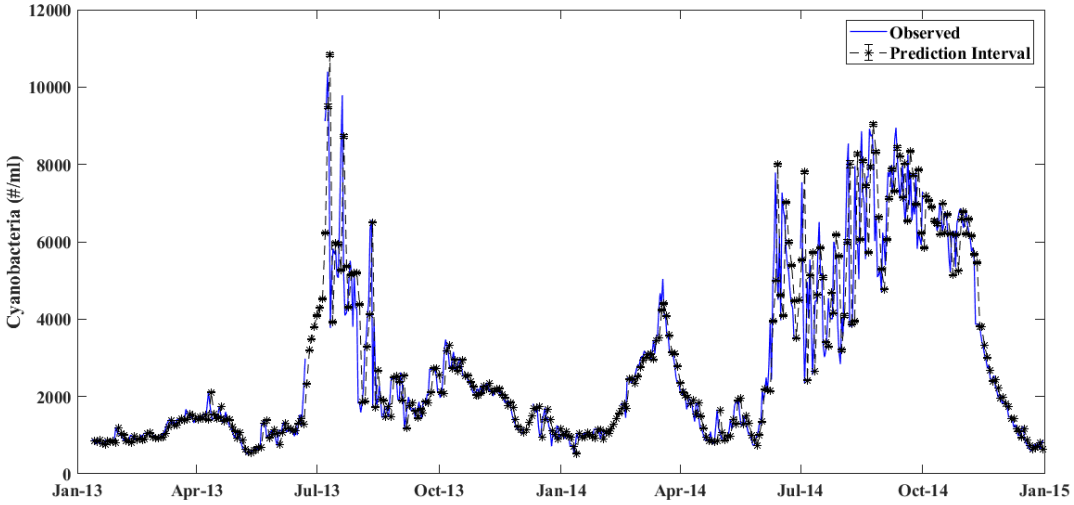
Different simulation time steps from 2 to 30 days were analyzed and used for model parameter calibration using Model 4. Calibration was based on data from 2013-2014 in Cheney Reservoir. Sets of optimized parameter values and corresponding statistics are summarized in Table 3-8. Depending on the time step, RMSE values range from 986 to 1502 cells/ml, and NSE values are from 0.61 to 0.80. Series of observed and simulated cyanobacteria concentration with different simulation time-steps (2 days, 7 days, 15 days, and 30 days) are shown in Figure 3.13 using calibrated parameters from Model 4.

Calibrated parameter values in Table 3-8 are within the acceptable range from the ones reported in the literature (Table 3-6). Model calibration for 2-day time-step simulation showed better statistics than for other intervals showing the highest NSE (0.801) and lowest RMSE (986). Model performance for a 30-day time interval was the lowest. Simulations for the 15-day interval were better than the results for the 7-day time step. It can be observed from Figure 3.13 that model

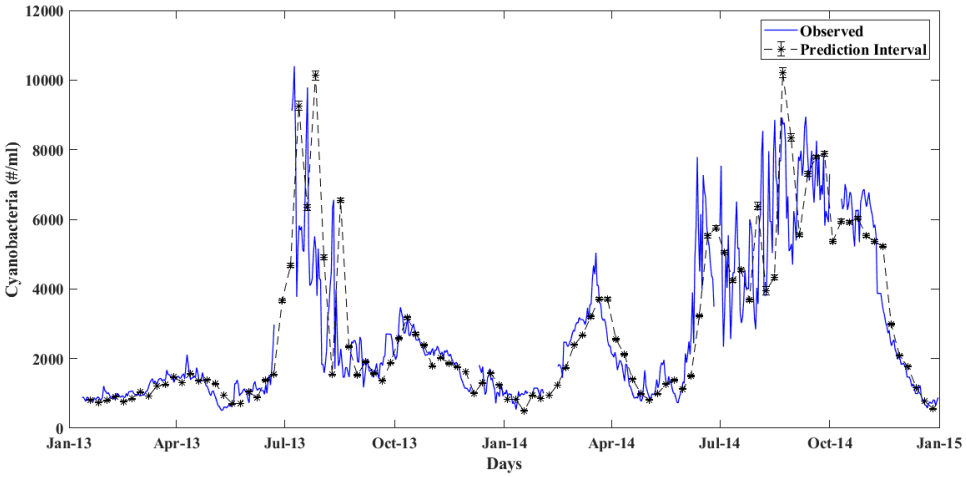
uncertainty increased while using a longer time step for model calibration and associated simulation.

Table 3-8 Optimized parameter values with 95% confidence interval (in brackets) and statistics using Model 4 for data averaged on a range of 2, 7, 15, and 30 days interval. ($\sigma = 0.27$, $\theta_\sigma = 1.08$, $\alpha = 0.01, \beta = 0.1$)

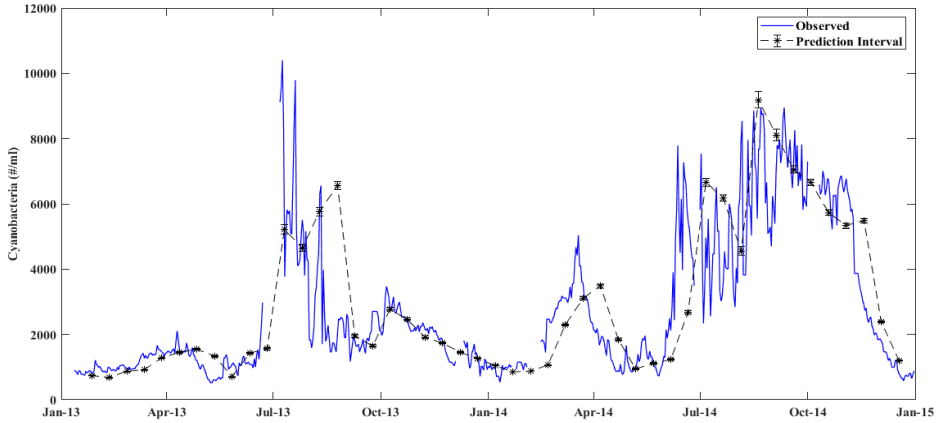
Parameters	μ	θ_μ	K_I	K_p	K_N	NSE (RMSE)
Days interval			**Cheney Reservoir**			
2	0.3064 (0.302, 0.311)	1.0824 (1.082, 1.083)	17.532 (15.869, 19.196)	1.119 (0.034, 2.203)	29.909 (19.620, 40.197)	0.8008 (986)
7	0.4025 (0.397, 0.408)	1.0759 (1.075, 1.076)	93.697 (90.061, 97.333)	1.00 (0.067, 1.933)	30.00 (20.837, 39.163)	0.6501 (1378)
15	0.3571 (0.353, 0.361)	1.0780 (1.0776, 1.0784)	49.4855 (47.278, 51.693)	4.2465 (3.313, 5.179)	29.995 (20.035, 39.955)	0.6832 (1301)
30	0.3665 (0.361, 0.373)	1.0784 (1.0778, 1.079)	43.5091 (40.439, 46.580)	10.00 (8.428, 11.572)	30.00 (14.75, 45.25)	0.6107 (1502)
Reference Value α=0.004-0.035, β=0.027-0.12	0.20-0.86	1.01-1.066	60-200	2-60	1-34	-



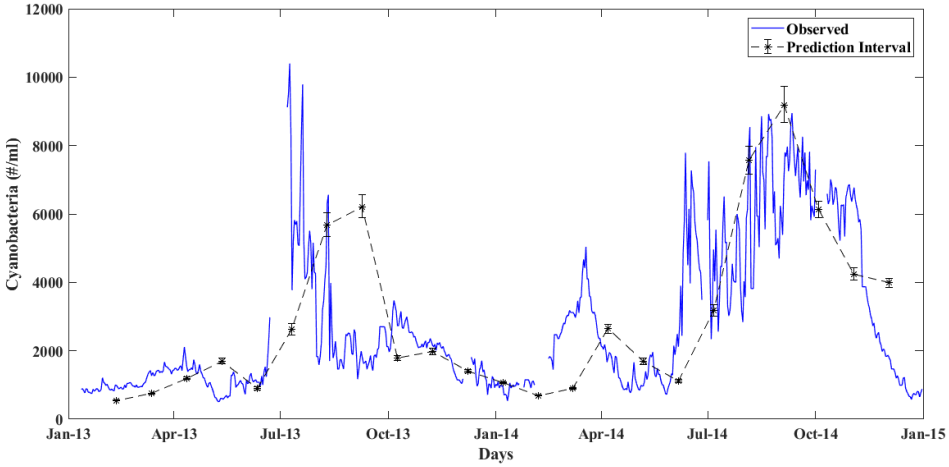
(a) 2 days averaged



(b) 7 days averaged



(c) 15 days averaged



(d) 30 days averaged

Figure 3.13 Cyanobacteria concentration outcomes using Model 4 optimized parameters calibrated from (a) 2 days averaged, (b) 7 days averaged, (c) 15 days averaged, and (d) 30

days averaged data sets for Cheney Reservoir. Simulation efficiency has a varied range for different intervals.

3.3.1.3 Environmental Factors and Growth Rate Curves

Cyanobacteria growth rates (observed and simulated) are plotted in Figure 3.14 and Figure 3.15 for Model 4 on the planes of temperature-nutrient (P or N) and temperature-irradiance. The values of the corresponding environmental variable are colored according to color bar ranges for each plot. Higher growth rates were observed when temperature, irradiation, and phosphorus are also higher in their values. Growth rates on the nitrogen planes did not have distinctive regression fit patterns like the ones seen for other environmental variables in the model.

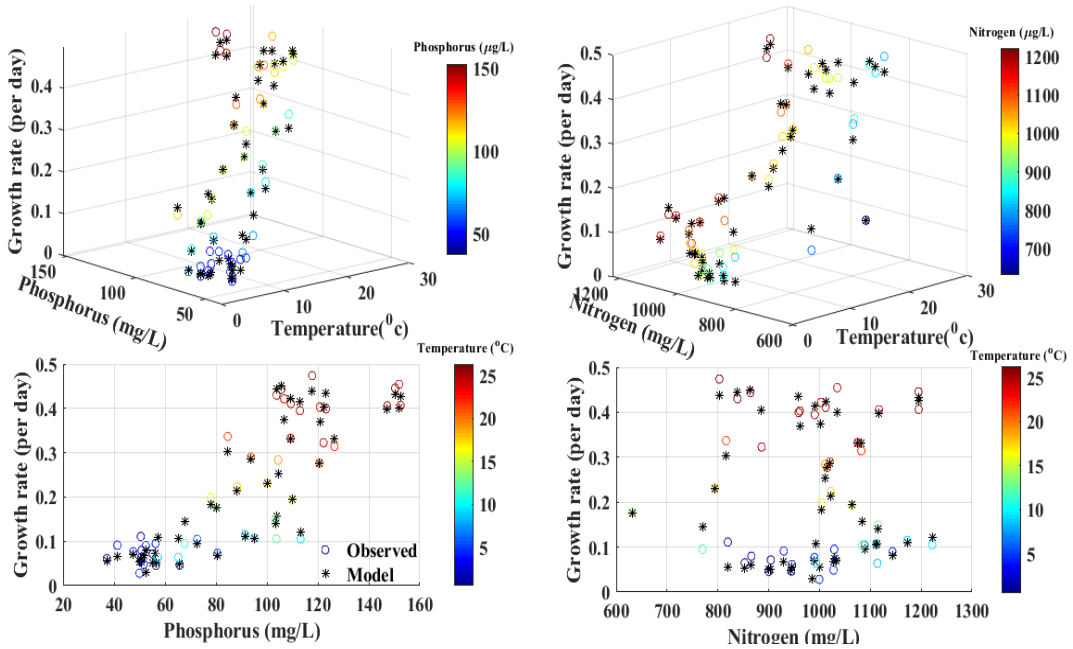


Figure 3.14 Observed and simulated growth rate of cyanobacteria on the planes of water temperature and nutrients using Model 4 in Cheney Reservoir. Here, Phosphorus (μg L^{-1}) and nitrogen (μg L^{-1}) refers to available phosphorus and available nitrogen for cyanobacteria.

Figure 3.15(a-b) reveals a periodic pattern of growth rate on the temperature-irradiation plane. It is observed that temperature can be seen as a primary factor for the growth of cyanobacteria.

During the increase in cyanobacteria (arrow points up), the growth rate starts to rise rapidly in the combination of higher irradiance and temperature. In Figure 3.15b, the growth rate increases from 0.1 day^{-1} to over 0.4 day^{-1} is sensitive to temperature increase from below 10 ^{o}C to over 20 ^{o}C, while irradiation remains close to 250 W/m^2. The growth rate rapidly increases once the temperature exceeds 15 ^{o}C with a slope of 0.02295 $day^{-1}/^{0}$C compared to 0.00577 $day^{-1}/^{0}$C for temperature below 10 ^{o}C (Figure 3.15c). This may be due to the combined interactions of temperature and other driving factors, such as light and nutrients, that create a favorable environment for cyanobacteria to grow. Multi-collinearity among temperature, irradiation, and phosphorus in Table 3-2 supports this behavior.

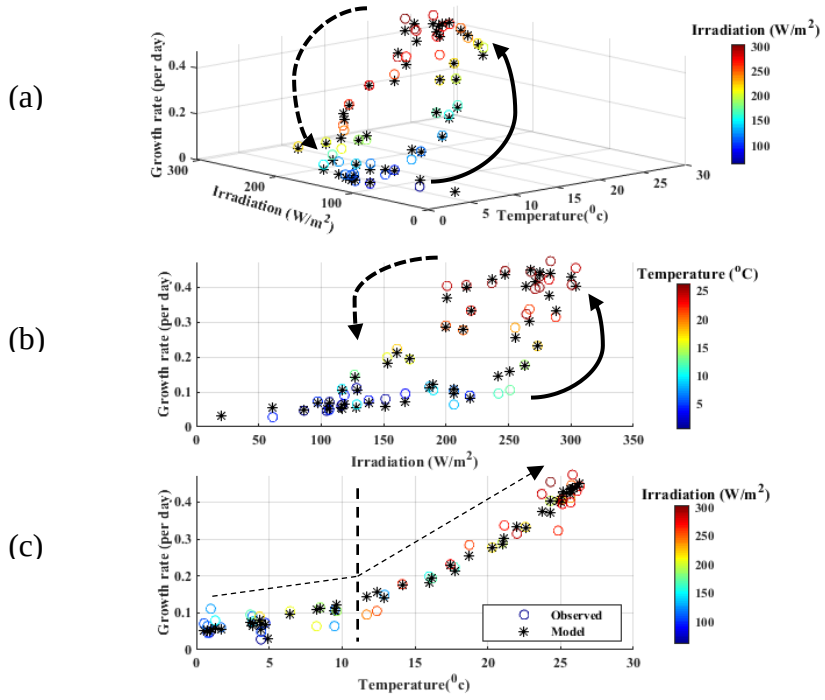


Figure 3.15 Observed and simulated growth rate for cyanobacteria on the planes of water temperature and irradiance using Model 4 in Cheney Reservoir. Growth rate curves on the temperature-irradiation plane show periodic patterns. Solid and dashed circular arrow in figure (a-b) indicate the rising and recession cycle of cyanobacteria respectively.

3.3.1.4 Qualitative model validation

For model validation, observed cyanobacteria concentration data were not available for other periods outside of the 2013-2014 years in Cheney Reservoir. Thus, phycocyanin concentration data was used for validation instead. The phycocyanin monitoring daily data was available for 2018 at the same sampling locations as data in 2013-2014. Phycocyanin is known to linearly correlate to cyanobacteria concentration according to Brient et al. (2008, p. 249) and Kong et al. (2014, p. 37) with a correlation coefficient of 0.997 for cyanobacteria group Anabaena, Planktothrix (Oscillatoriales), and Microcystis. Therefore, we used phycocyanin data for validation of the cyanobacteria growth model in the Cheney Reservoir. Based on these studies, a linear relationship with coefficient of 0.001 ($y = 0.001x$) was used for estimating cyanobacteria (y, cells/ml) from phycocyanin (x, µg/L) observations.

From Table 3-7, calibrated parameters based on Model 2 were used during cyanobacteria forecasting. 15-day average water temperature and irradiation were used as model input variables. Time-series of observed and simulated growth rates for 2018 and associated patterns in the temperature-irradiance plane are shown in Figure 3.16 and Figure 3.17. Cyanobacteria cell concentration simulation and comparison with phycocyanin trends for the year 2018 are shown in Figure 3.18a. A comparison of simulated and observed growth rates showed good model validation with NSE (0.56) and RMSE (1459) (Figure 3.18b). We note that the values of the parameters were taken from calibrated Model 2 and were not specifically adjusted for the year 2018.

Cyanobacteria contain phycocyanin pigments, and this is the reason for their bluish appearance. Phycocyanin has a significant correlation with cyanobacteria cell concentrations and can be used to develop cyanobacteria monitoring tools (Brient et al. (2008), Kong et al. (2014)). So, we have used phycocyanin monitoring data from the Cheney reservoir to validate our

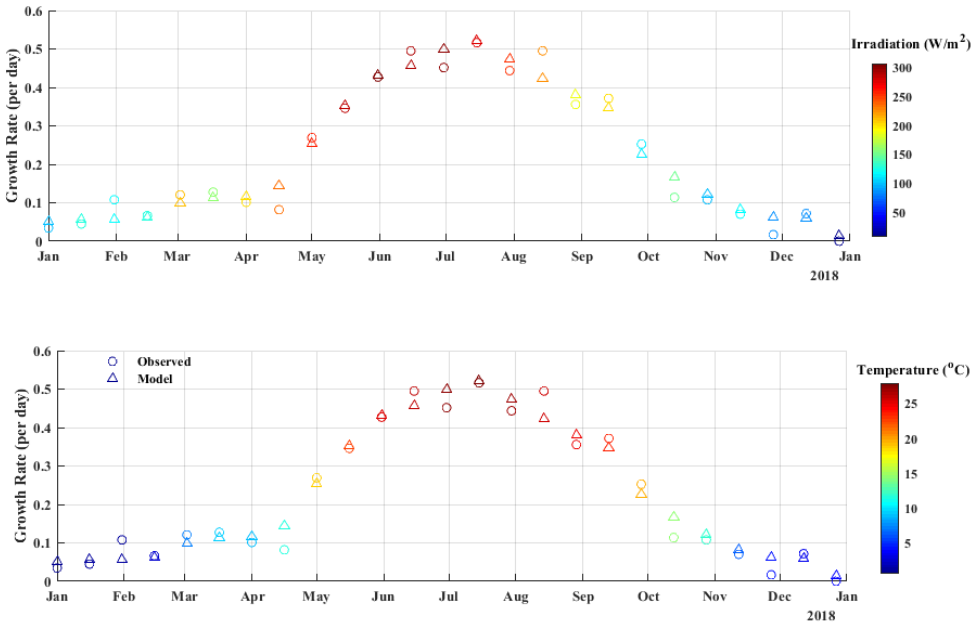


Figure 3.16 Comparison of cyanobacteria growth rate based on measured phycocyanin during the year 2018.

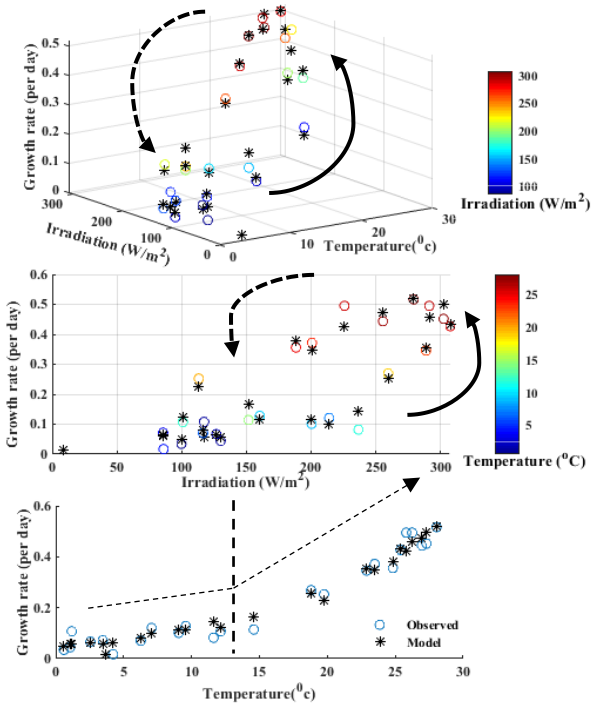
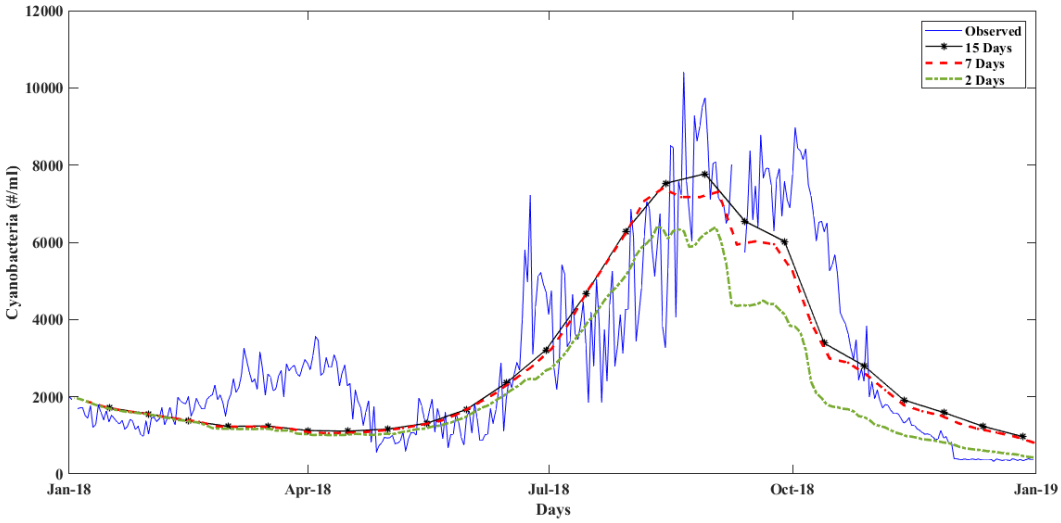
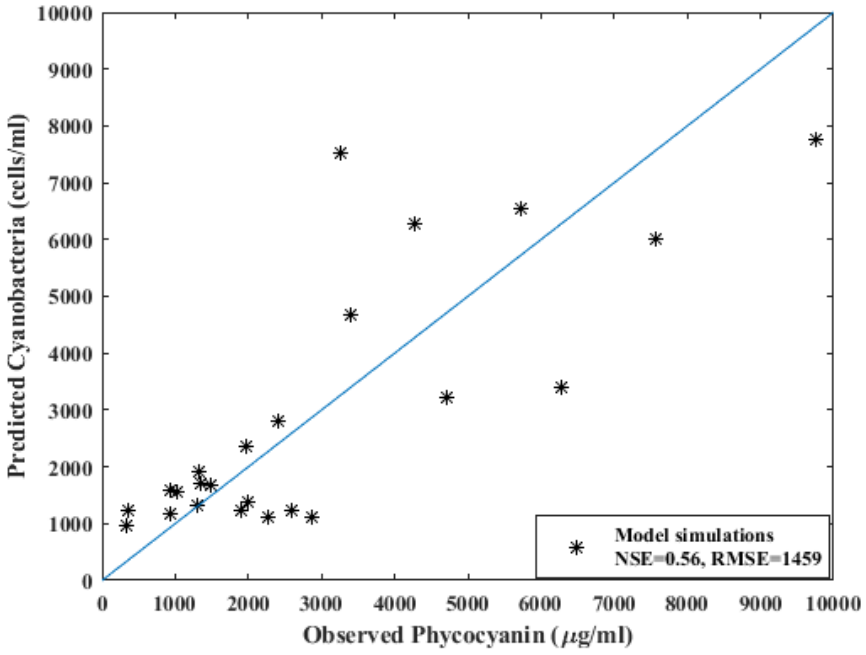


Figure 3.17 Observed and simulated growth rate on the planes of water temperature and irradiance using Model 2 in Cheney Reservoir based on data from 2018.



(a)



(b)

Figure 3.18 Model simulations for forecasting of cyanobacteria cell concentration for 2018 in the Cheney reservoir using Model 2 (RMSE=1459, NSE=0.56 for 15 days) (a-b). The continuous blue line in (a) indicates estimated cyanobacteria concentration based on observed daily phycocyanin in µg/ml. Black, red and green lines indicate model outcomes using optimized parameters from Model 2 (Table 3-7) and updating environmental variables (T and I) at every 15, 7, and 2 days intervals respectively. (b) Model performance when T and I were updated every 15 days interval.

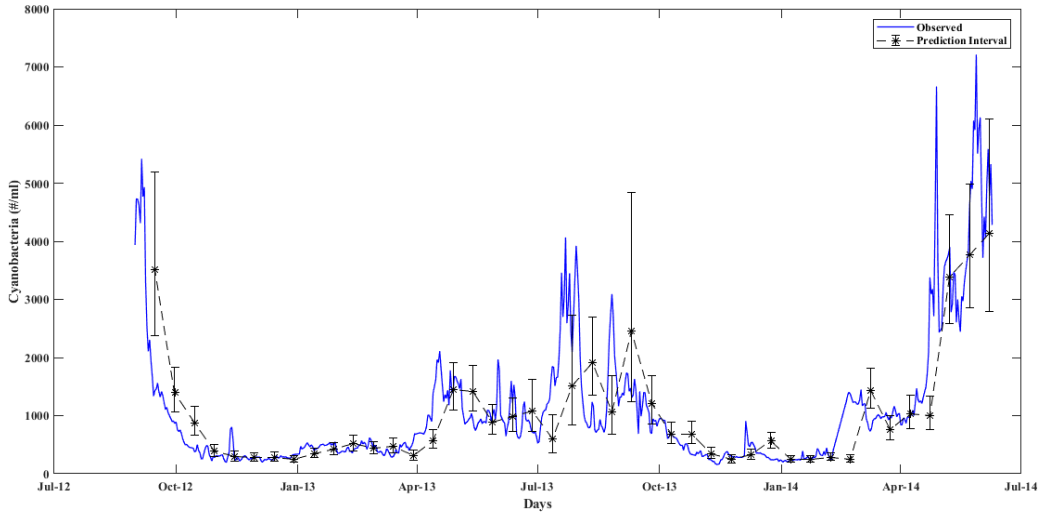
developed model. Model 2 was used for validation purpose due to the availability of continuous data for water temperature and irradiation. Phycocyanin monitoring data from 2016-2018 was available. Data from 2018 had relatively low water inflow and hence smaller turbulence in the water surface comparing to 2016-2017. Since the developed model assumed a reservoir as a constantly stirring tank with a well-mixed condition, it justifies using the 2018 dataset. It can be noted that there was no continuous phycocyanin monitoring data in 2019-2020.

3.3.2 Model Simulation for Kansas River (Wamego)

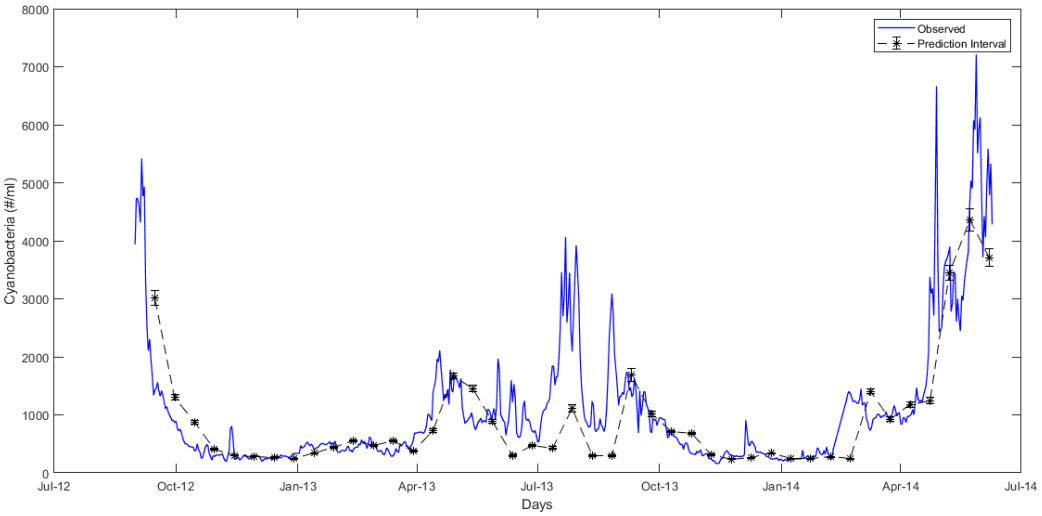
Model 1 and Model 2 were applied to the site in Kansas River at Wamego. The models were calibrated using the approach presented in section 3.2.4 and optimized parameter values are listed in Table 3-9. These parameter values are within the range of reference values from previous studies. Model simulation results are shown in Figure 3.19 and Figure 3.20. The values were simulated for 15-day intervals and each interval calculation used observed cyanobacteria concentration at the beginning of the interval and 15-day interval average values of temperature and irradiation. The results showed that the model was able to simulate cyanobacteria cell concentrations for short-term prediction with RMSE in the range of 735-758 cells/l and NSE in the range of 0.39-0.45.

Table 3-9 Model parameters with optimized values with 95% confidence interval (in brackets) for two simulation models in Kansas River at Wamego, KS ($\sigma = 0.27$, $\theta_{\sigma} = 1.08$)

Parameters	μ_m (day^{-1})	θ_{μ}	K_I (W/m^2)	NSE (RMSE)
Kansas River (Wamego)				
Model 1 (T)	0.2682 (0.254,0.283)	1.077 (1.068, 1.086)		0.4530 (735)
Model 2 (T, I)	0.3276 (0.321, 0.335)	1.0645 (1.0636, 1.0654)	32.60 (27.46, 37.75)	0.3946 (758)
Reference Value	0.20-0.86	1.01-1.066	60-200	-

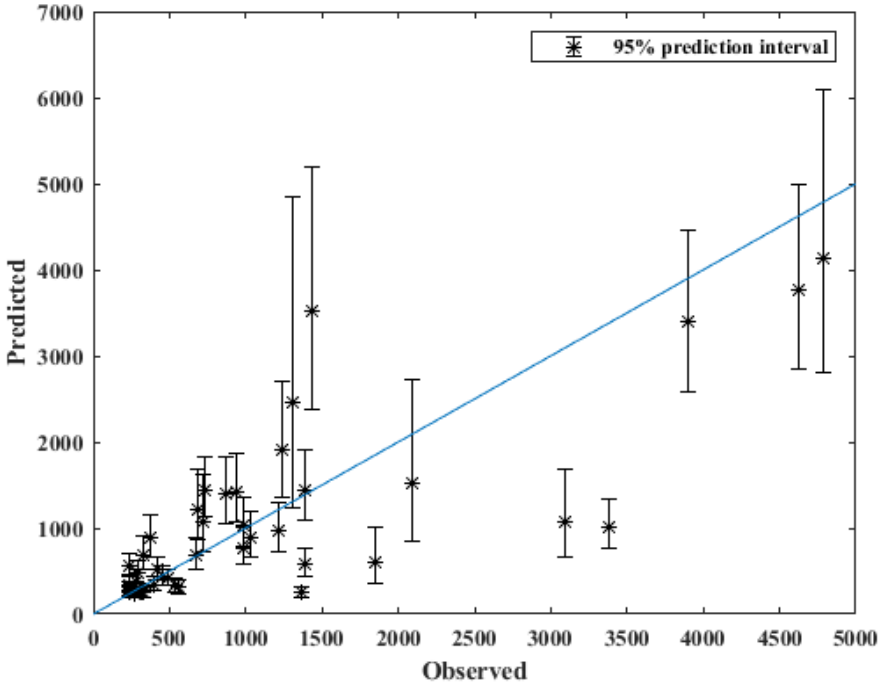


(a) Model 1

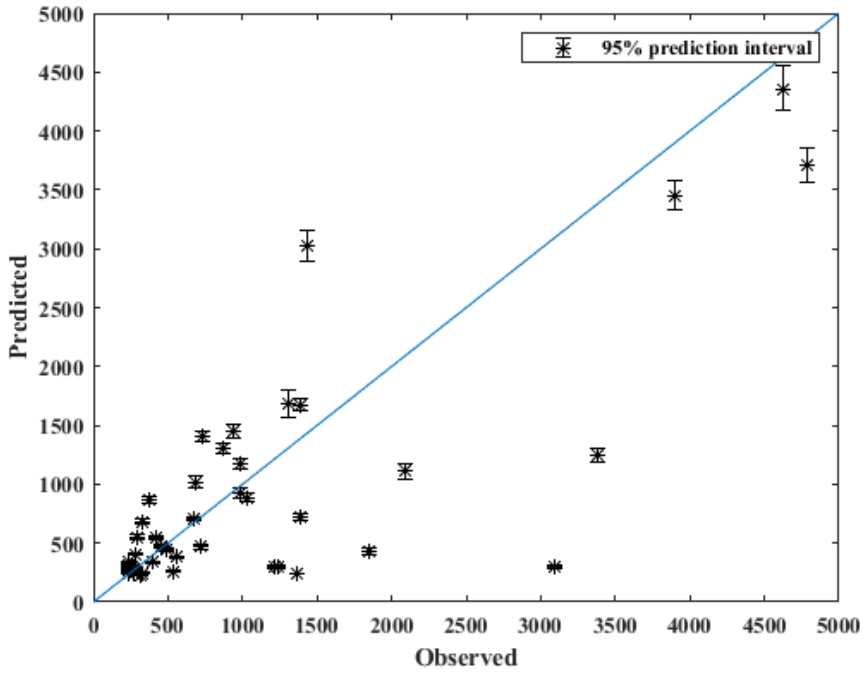


(b) Model 2

Figure 3.19 Model simulations for 15-day backcasting of cyanobacteria cell concentration in Kansas River, Wamego using (a) Model 1 (RMSE=735, NSE=0.4530), and (b) Model 2 (RMSE=758, NSE=0.3946). The continuous blue line indicates observed daily cyanobacteria concentrations in cells/ml, and the black dashed line indicates model outcomes using optimized parameters and environmental variables.



(a) Model 1



(b) Model 2

Figure 3.20 Simulated vs. observed cyanobacteria concentration (#/ml) for CyanoHABs forecasting from (a) Model 1 and (b) Model 2 in Kansas River, Wamego.

Figure 3.21 shows a periodic pattern of the growth rate on the temperature-irradiation plane. This pattern is very similar to that found for Cheney Reservoir in Section 3.3.1.3. The temperature effect on the growth rate is found prominent in this study site also. It is also observed that the growth rate rapidly increases with a slope of 0.02692 $day^{-1}/^{0}C$ when the temperature exceeds 15 ^{o}C compared to 0.0070 $day^{-1}/^{0}C$ for the temperature below 10 ^{o}C.

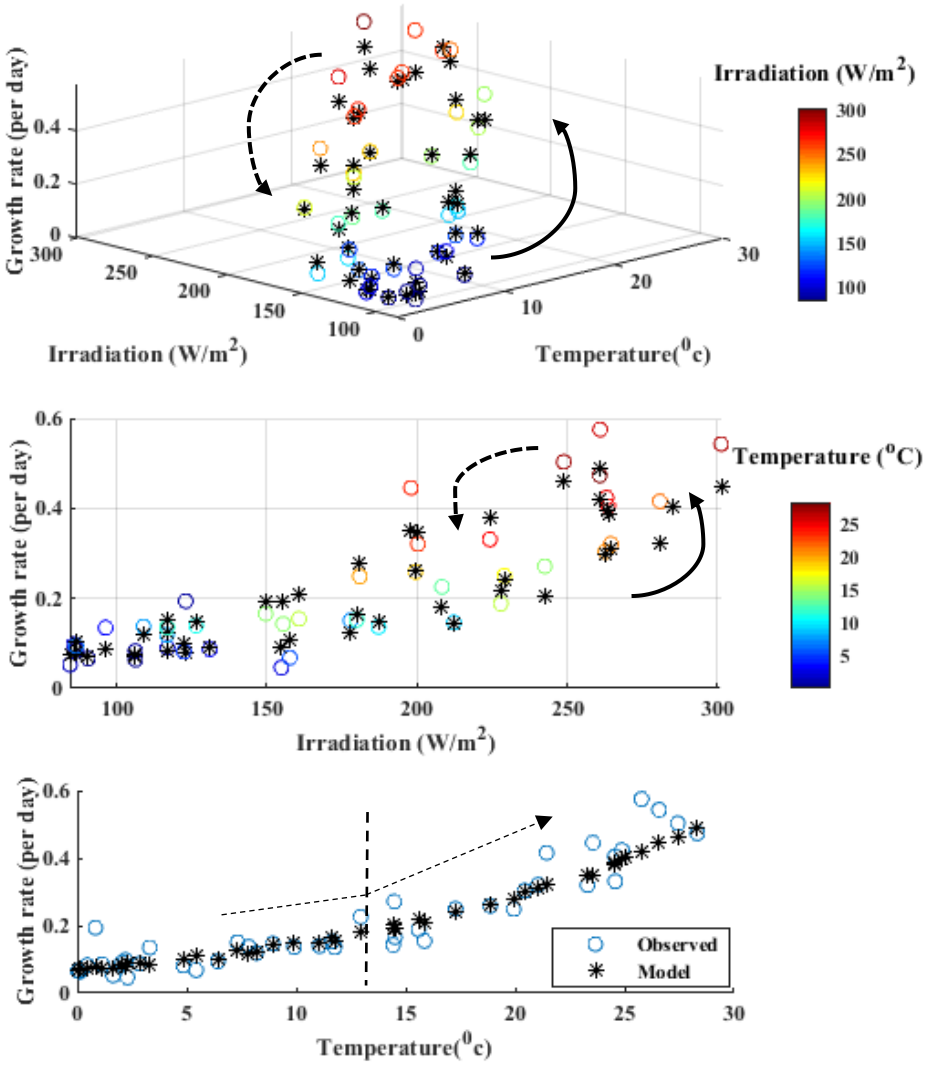


Figure 3.21 Observed and simulated growth rate for cyanobacteria on the planes of water temperature and irradiance using Model 2 in Kansas River, Wamego.

3.3.3 Summary Discussions

Short-term modeling in data-scarce eutrophic lakes:

Four cyanobacteria growth models were developed based on four main environmental factors: water temperature, irradiation, phosphorous concentration, and nitrogen concentration. All models were compared with the observed dataset in Cheney Reservoir and Kansas River at Wamego. The statistics calculated for Model 1, Model 2, and Model 4 in Cheney Reservoir showed that reduced input Models 1 and 2 were comparable to full input Model 4 (sec. 3.3.1.1). This allowed us to use the reduced models (1 and 2) for conditions where nutrient concentrations were not available from observed datasets. We remind that Model 1 requires only temperature as a driving factor, whereas Model 2 takes temperature and irradiance.

Cheney Reservoir is a eutrophic lake (Wang et al., 2003), and cyanobacteria growth rate is not limited by only nutrients (i.e. $\varphi_P \sim 1$ and $\varphi_N \sim 1$ in equation [3.12c-3.12d]). Therefore, the eutrophic nature of the Cheney Reservoir can be considered an underlying factor for high efficiency of the reduced models (Model 1 and Model 2) when compared to the full model (Model 4). The efficiency of temperature-dependent Model 1 can also be attributed to multicollinearity among temperature, phosphorus (0.84), and irradiation (0.77). Thus, changes in temperature can implicitly capture the effect of the changes in phosphorus and irradiation.

More than half of the lakes in the United States and specifically in the Midwest are eutrophic. Continuous monitoring is important for understanding the dynamics in seasonal and spatial nutrient variability within a lake. Unfortunately, it is unavailable in many of those lakes. Our analysis showed that nutrient limitations can be ignored for cyanobacteria growth estimates in eutrophic water bodies, and using models 1 and 2 with the assumption of $\varphi_P \sim 1$ and $\varphi_N \sim 1$ can be appropriate.

Effect of data sampling frequency on the model:

Four time-intervals for model simulation were tested for short-term forecasting in Cheney Reservoir and Kansas River at Wamego (2-day, 7-day, 15-day, and 30-day). Model calibration and simulation efficiency statistics presented in Table 3-8 indicate that the 2-day time interval had better performance than larger time-steps. However, updating cyanobacteria concentration values, as well as other physical variables that are needed as initial condition for model simulation, can be costly and time-consuming due to expensive monitoring setups and laborious resources. Since 15-day interval statistics was on par with 2-day statistics, updating every 15 days instead of 2 days can be seen as a viable option for forecasting simulations.

Growth rate patterns in natural environments:

In the literature and from previous studies, available growth rate curves can be found based on experimental data obtained in a controlled laboratory environment and for single species of phytoplankton. Analysis from sec. 3.3.1.3 revealed that multiple factors and phytoplankton species can constrain growth rate patterns in a natural lake environment. In this study, in nutrient-rich water bodies, the effects of P and N were less prominent on HAB prediction than the impact of temperature and irradiation (Figure 3.14, Figure 3.15). As a result, under eutrophic conditions cyanobacteria growth rates can be predicted by considering fewer driving factors and primarily water temperature and sunlight. Growth rate patterns are more affected by nutrients during periods when nutrients are not abundant in waterbodies.

Long-term predictions and lake management

Nutrient based models, Model 3 (T, I, P) and Model 4 (T, I, P, N) performed slightly better than the reduced input models for long-term and short-term predictions (Figure 3.12a-b). Model 3 and Model 4 followed the pattern of cyanobacteria change during two bloom periods well.

However, it had a lower agreement in winter-spring seasons. For January-April, when temperatures are moderate, the model may need to have a different set of optimized parameter values.

3.4 Conclusions

The developed mechanistic modeling framework for cyanobacteria growth in well-mixed water bodies was capable of making short-term predictions of cyanobacteria concentrations. The framework was tested in two water bodies in Kansas and provided adequate results. Four cyanobacteria growth models were introduced. Each contained a different number of physical variables. The factors of water temperature, irradiation, and phosphorous and nitrogen concentrations were used as driving environmental factors. Nutrient based models (Model 3 and 4) performed better than reduced models (Model 1 and 2) for short-term and long-term cyanobacteria forecasting. However, performance of the reduced models can be significant for short-term (day to bi-weekly) cyanobacteria concentration forecasting in nutrient-enriched eutrophic conditions. Our analysis revealed that 15-day interval data monitoring can be optimum for model application purposes.

Our data analysis found a significant correlation of cyanobacteria cell concentrations with temperature, sunlight, reservoir volume, and phosphorus concentrations. Observed growth rate pattern of cyanobacteria impacted from combined interactions of temperature, irradiance and nutrient in the natural environment can help HAB management. In the future, HAB occurrences can be negatively impacted due to more extended drought periods with warmer weather, stagnant water, nutrients influx, and sunlight. Implementation of conservation practices on agricultural fields can be practiced by reducing the amount of nutrients reaching streams and lakes, and as a result, controlling cyanobacteria growth rate and limiting the frequency of the blooms.

The mechanistic modeling framework in this study assumed the presence of other competitor phytoplankton groups of cyanobacteria and zooplankton grazing negligible. Other external factors like surge inflow, airspeed, turbulence, viral lysis, nutrients uptake, and internal cell biology that were not considered into the model may have effects on the bloom formation. These external factors can affect both the growth and decay of cyanobacteria by carrying new species into the lake and changing the aquatic environment. Future study incorporating some of those factors into the modeling framework can improve our understanding of HAB dynamics.

Chapter 4 Harmful Algal Bloom: Stochastic Modeling

4.1 Background

Modeling of Harmful Algal Blooms (HAB) in aquatic systems usually focuses on simulating the dynamics of cyanobacteria concentration in the lake considering biophysical processes (bacteria growth and decay), impacts of environmental factors, and other system dynamics like water flow and transport of nutrients (Chapter 3). A mathematical model is a simplified representation of a complex natural system in which the processes are triggered or suppressed by biological and/or physical activities (Wang, 2011, pp. 2-8). Variability in the representation of driving processes contributes to uncertainty in model formulation. In case of HABs, uncertainty is manifested in the errors in collected data for physical variables, fluctuations in values of model parameters within a single modeling time-step, specific representation of underlying biophysical processes, and others. Uncertainty quantification in ecological systems involves parameters estimation and representation of complex interactions during model application. Understanding an overall model uncertainty and the stochastic process can help probabilistic risk assessment of HAB events. Developing a stochastic modeling framework to quantify HAB forecasting uncertainty can also be necessary for scenario development in the long-term projected climate change scenarios (Ralston and Moore, 2020, pp. 7-10).

In this chapter, we present a framework of stochastic modeling of cyanobacteria dynamics based on the assumption that influent nutrient concentrations can be affected by the fluctuations in incoming stream discharge. Use of the Fokker-Planck equation ranges from describing the Brownian motion of particles, noise in the quantum optics, population growth dynamics, reservoir storage volume change to algal modeling (Risken, 1989; Soboleva and Pleasants, 2003; Zielinski, 1984; Huang et al., 2008; Wang et al., 2013). The Fokker-Planck equation is implemented that

provides probabilistic density functions (PDF) of the algal bloom concentrations. This enables us to understand the uncertainty of the non-linear system with many unknown parameters. The Fokker-Planck equation is also used to understand the effect of variability of initial concentrations on the model uncertainty. Temporal evaluations of the concentrations PDF can quantify the likelihood of certain events that can cross alert levels of HAB based on reference guidelines, thereby taking preparatory management actions.

4.2 Problem Statement

Nutrients (P and N), water temperature, and sunlight irradiation are considered major driving environmental factors for process-based phytoplankton modeling in the aquatic system as described in Chapter 3. Cyanobacteria growth model development assumes continuously stirred tank reactor (CSTR) and mass balance equations of different algal groups, such as diatoms, green algae, cyanobacteria, and other minor groups. A generalized ordinary differential equation (4.1) is used to model phytoplankton concentrations A_i (i=1, 2, … ,n) with *n* algae groups in the aquatic system. The model considers bacteria natural growth (GR_i), non-predatory loss (NL_i), grazing loss (GL_i) and washout (WO) as physical rates of change in the natural phytoplankton system. The resulting system of equations is:

$$\frac{dA_i}{dt} = (GR_i - NL_i - GL_i - WO)A_i \tag{4.1}$$

where,

$$GR_i = \mu_{m_i}\theta_{\mu_i}^{T-T_{Ref}}\frac{P}{K_{P_i}+P}\frac{N}{K_{N_i}+N}\frac{I}{K_{I_i}+I} \tag{4.1a}$$

$$NL_i = \sigma_{m_i}\theta_{\sigma_i}^{T-T_{Ref}} \tag{4.1b}$$

$$WO = \frac{q_{out}}{V} \tag{4.1c}$$

$$P = P_{tot} - \sum_{i=1}^{n}\alpha_i A_i \text{ and } N = N_{tot} - \sum_{i=1}^{n}\beta_i A_i \tag{4.1d}$$

GR_i and NL_i are the rates that depend on water temperature and are modeled using the theta model in equations (4.1a), (4.1b). Phytoplankton growth affected by nutrients and irradiance is modeled using the Monod kinetics, which assumes an approximately linear increase of growth rate up to half-saturation coefficients (amounts where growth is limited to half of the optimum condition) and then gradual increase towards the maximum in nutrients or light abundance regions (equation (4.1a)). Available nutrients (*P* and *N*) for the phytoplankton groups in the system are dependent on reservoir nutrients concentrations (P_{tot} andN_{tot}) and internal nutrients concentrations within the algae groups. Internal nutrient contents (α_i and β_i) are used to model available nutrients for the phytoplankton which is described in equation (4.1d). Zooplankton grazing rate depends on the amount of zooplankton presence and type of phytoplankton groups. WO depends on the outflow rate $\mathrm{q_{out}}$ of the system and reservoir storage volume *V*.

Here, we use a simplified form of equation (4.1) that models only one group of phytoplankton (i.e. cyanobacteria) for implementing stochastic framework. The reduced model for only cyanobacteria concentration (*C*) and associated terms are stated in equation (4.2) in which zooplankton grazing effect on the cyanobacteria is assumed minimal.

$$\frac{dC}{dt} = (GR - NL - WO)C \tag{4.2}$$

where,

$$GR = \mu_m \theta_\mu^{T-T_{Ref}} \frac{P}{K_P + P} \frac{N}{K_N + N} \frac{I}{K_I + I} \tag{4.2a}$$

$$NL = \sigma_m \theta_\sigma^{T-T_{Ref}} \tag{4.2b}$$

$$WO = \frac{q_{out}}{V} \tag{4.2c}$$

$$P = P_{tot} - \alpha C \text{ and } N = N_{tot} - \beta C \tag{4.2d}$$

We can observe from equation (4.2) that it has several parameters to be optimized from experimental observations or can be found from earlier studies. These parameters are highly uncertain that suggests quantifying the uncertainty of the modeling system. Hence, we aim to understand the stochastic details of the HAB dynamics and quantify PDF of the cyanobacteria

concentration, *W (ξ, t)*. We used ξ as the deterministic outcomes of random population concentrations of cyanobacteria. During the stochastic analysis water temperature, irradiation, outflow, and reservoir volume are considered steady with time. Nutrients fluctuations are taken into account to understand the impact of the gradual nutrient loading phenomenon into the lake following the steps described in Wang et al. (2013, pp. 100-102) and further extending for dimensionless analysis.

4.3 Methods

Nutrients are one of the triggering factors of algal blooms that fluctuate temporally. Nutrients loading into the lake depends on runoff events. A simplified lake nutrients model in equation (4.3) is used to simulate temporal fluctuations of total nutrients concentrations (P_{tot} and N_{tot}) in the reservoir as impacted from inflow volume or cumulative runoff *Q(t)*, and inflow nutrients concentrations (c_P and c_N for phosphorus and nitrogen influx concentrations respectively). P_0 and N_0 denote initial phosphorus and nitrogen concentrations in the reservoir respectively.

$$P_{tot} = P_0 + c_P \frac{Q}{V}, \qquad N_{tot} = N_0 + c_N \frac{Q}{V} \tag{4.3}$$

We can represent cumulative runoff *Q(t)* as its ensemble average ($\bar{Q}$) and zero-mean fluctuations (Q') based on Reynolds decomposition as in equation (4.4).

$$Q(t) = \bar{Q} + Q' \tag{4.4}$$

We use Taylor expansion series around $\bar{Q}$ to express the random growth rate as

$$GR = GR(\bar{Q}) + \frac{\partial GR}{\partial Q}(\bar{Q})Q' + \frac{\partial^2 GR}{\partial Q^2}(\bar{Q})Q'^2 + \mathcal{O}(Q'^3) \tag{4.5}$$

Replacing the first two terms from equation (4.5) to (4.2) and rearranging it yields

$$\frac{dC}{dt} = (GR(\bar{Q}) - NL - WO)C + \frac{\partial GR}{\partial Q}(\bar{Q})CQ'(t) \tag{4.6}$$

Equation (4.6) can be transformed into its dimensionless form (4.7) in the form of a Langevin equation with white noise $Q'^{*}(t)$:

$$\frac{dC^*}{dt^*} = h^*(C^*, t^*) + g^*(C^*, t^*)Q'^{*}(t) \tag{4.7}$$

Where,

$$h^*(C^*, t^*) = (GR(\bar{Q}) - NL - WO)\frac{C^*}{\mu_m} \tag{4.7a}$$

$$g^*(C^*, t^*) = \left(\frac{\partial GR}{\partial Q}C^*\right)\frac{V}{\mu_m} \tag{4.7b}$$

$$C^* = \frac{C}{C_0}, \quad t^* = t\mu_m \text{ and } Q'^{*} = \frac{Q'}{V} \tag{4.7c}$$

Normalizing factors C_0 and μ_m in equation (4.7c) represent initial concentrations and maximum growth rate of cyanobacteria, respectively. Following the standard procedure described in Risken (1989, pp. 44-50), one can obtain the Fokker-Planck equation (advection-dispersion equation) for PDF of the cyanobacteria concentrations W (ξ^*,t^*) in dimensionless time-concentration space as:

$$\frac{\partial W(\xi^*, t^*))}{\partial t^*} = \left[-\frac{\partial}{\partial \xi^*}D_1^*(\xi^*, t^*) + \frac{\partial^2}{\partial {\xi^*}^2}D_2^*(\xi^*, t^*)\right]W(\xi^*, t^*) \tag{4.8}$$

$$D_1^*(\xi^*, t^*) = h^*(\xi^*, t^*) + g^*(\xi^*, t^*)\frac{\partial g^*(\xi^*, t^*)}{\partial \xi^*} \tag{4.8a}$$

$$D_2^*(\xi^*, t^*) = {g^*}^2(\xi^*, t^*) \tag{4.8b}$$

Substituting the terms of equations (4.7a-b) into equations (4.8a-b) yields the following forms of convection D_1 and dispersion D_2 coefficients:

$$D_1^*(\xi^*, t^*) = (GR(\bar{Q}) - NL - WO)\frac{\xi}{C_0\mu_m} + \left(\frac{\partial^2 GR(\bar{Q})}{\partial Q \partial \xi}\xi + \frac{\partial GR(\bar{Q})}{\partial Q}\right)\frac{\partial GR(\bar{Q})}{\partial Q}\xi\frac{V^2}{{C_0}{\mu_m}^2} \tag{4.9a}$$

$$D_2^*(\xi^*, t^*) = \left(\frac{\partial GR}{\partial Q}\xi\right)^2\frac{V^2}{{C_0}^2{\mu_m}^2} \tag{4.9b}$$

This $W(\xi^*, t^*)$ can be transformed into a PDF of the cyanobacteria concentration $W(\xi, t)$ in the original scale where $\xi^* = \frac{\xi}{c_0}$.

4.4 Results and Discussions

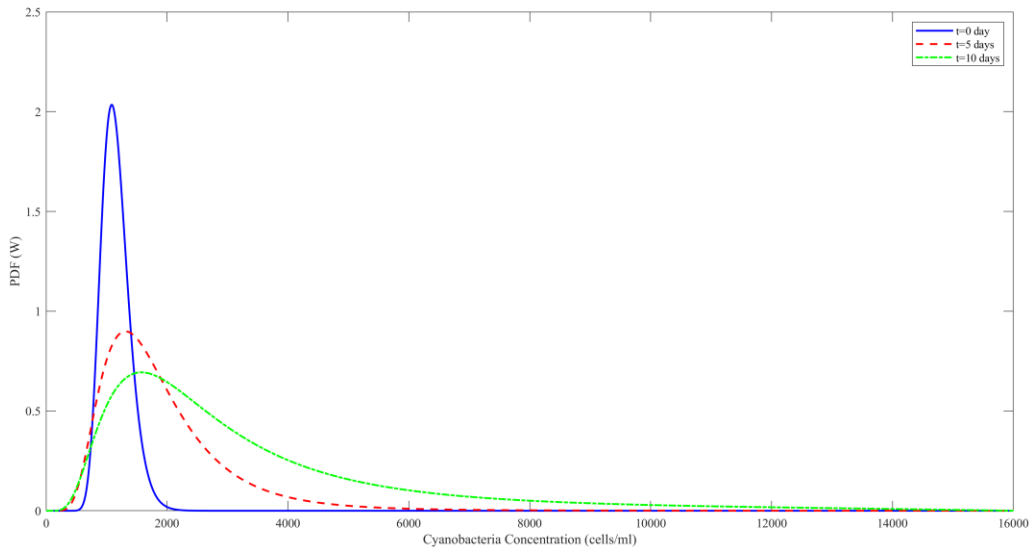


Figure 4.1 Evolution of the cyanobacteria concentration PDF W (ξ, t) along with time at t=0, 5 and 10 days

We analyzed data from Cheney Reservoir and used optimized values of driving parameters derived in Chapter 3 for the stochastic analysis. A lognormal distribution א (1, 0.2) with a mean concentration of 1 and a standard deviation of 0.2 is used as the initial PDF profile for simulating $W(\xi^*,t^*)$ in the dimensionless concentration and time-space. Model simulations for $W(\xi^*,t^*)$ are performed, and the results are presented in the form of dimensional $W(\xi,t)$ based on procedures described in the previous section.

Stochastic model outcomes for $W(\xi,t)$ for initial time and time of 5 and 10 days are shown in Figure 4.1. Influent nutrients coming with runoff volume along other environmental factors (water temperature and light) trigger the overall growth that results in general increase of cyanobacteria in the system from an initial mean concentration PDF of 1130 cells/ml to PDF of

about 3000 cells/ml. A gradual change of PDF with time can be observed as the result of the advection-diffusion processes defined by the coefficients D_1^* and D_2^* that were impacted by the changes in driving factors and model parameters. The temporal evolution of $W(\xi,t)$ reveals enlargement of distribution tails over time that indicates propagation of uncertainty of cyanobacteria forecast with time.

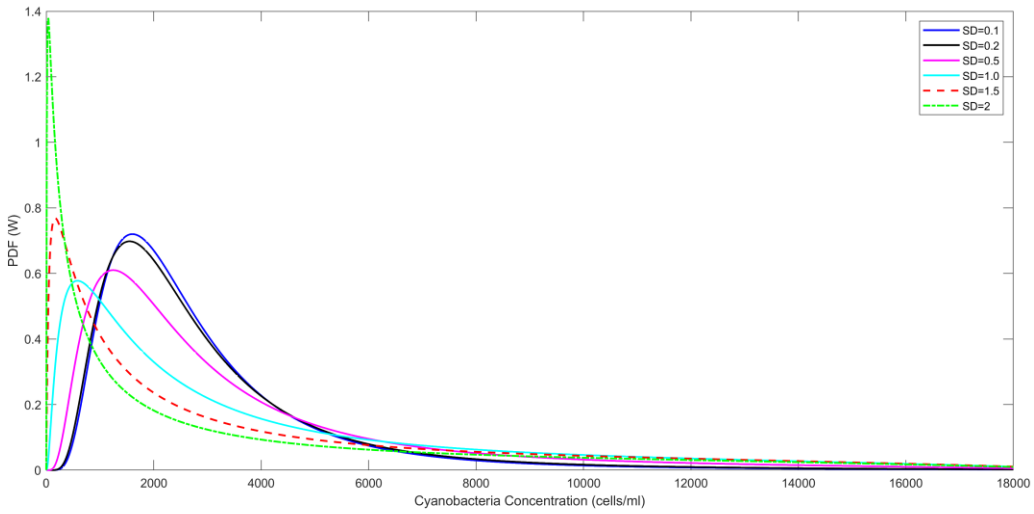


Figure 4.2 Comparison of the cyanobacteria concentration PDF W (ξ,t) at t=10 days considering uncertainty of the initial concentration profiles

The effect of the initial concentration uncertainty on cyanobacteria concentration PDF $W(\xi,t)$ observed in Figure 4.2 shows $W(\xi,t)$ for six different standard deviation (SD) at t=10 days. The distribution profiles at t=10 days indicate higher uncertainty (longer distribution tails) with an increased level of uncertainty at the initial concentration profiles (higher SD). This justifies the significance of the initial concentration impact on the uncertainty of the forecast. The difference in diffusion (peak of PDFs) is the outcome of different levels of the peak in the initial lognormal profile of the PDFs due to different SD.

4.5 Conclusions

In this section we implemented an uncertainty quantification methodology based on the Fokker-Planck equation (deterministic equation) and developed a solution that represents HAB concentration at different times in the form of PDF. PDF of cyanobacteria concentration quantifies the uncertainty and stochasticity of the HAB dynamics in a lake system. Analysis in the Cheney Reservoir based on this approach provides temporal snapshots of cyanobacteria concentration PDF to understand probabilistic risks in cyanobacteria propagation. A significant effect of uncertainty in initial concentration on the resulting uncertainty in cyanobacteria concentration forecasting is observed from the analysis. This suggests that facilitating a continuous lake monitoring system with more frequent updates to physical parameters can provide higher certainty in setting initial concentration and reduce uncertainty in cyanobacteria propagation. Further analysis is needed to quantify the impact of varying water temperature, nutrients, and irradiation on cyanobacteria concentration PDF.

Chapter 5 Summary and Future Research

5.1 Findings

This study investigated CyanoHAB dynamics from both mechanistic and stochastic modeling perspectives. Modeling HABs in a natural eco-system depends on multi-level processes and interactions among various environmental factors. Correlation analysis was conducted to understand the impact of environmental variables on the cyanobacteria concentration and was based on available monitoring data in Cheney Reservoir and Kansas River at Wamego. Cyanobacteria was found to significantly correlate with water temperature, irradiation, and phosphorus concentration for both study sites. In addition, turbidity and chlorophyll-a were found to correlate with cyanobacteria concentration at the Kansas River site. The opposite relationship between cyanobacteria and dissolved oxygen was observed at both sites. This evidence justifies depletion of water quality and points to ecological concerns at both studied sites as a result of abundance of cyanobacteria in waterbodies.

A process-based mechanistic modeling approach was implemented for both sites to facilitate a short-term and long-term cyanobacteria forecasting tool. Four different types of the mechanistic model were examined to assess its suitability for cyanobacteria prediction based on available data in 2013-2014 at the Cheney Reservoir site. Model parametrization was performed for all four models, followed by model calibration that revealed that Model 1 (T-based) and Model 2 (T, I- based) could be applicable for short-term (day to bi-weekly) forecasting, whereas Model 3 (T, I, P- based) and Model 4 (T, I, P, N- based) could better be used for long-term forecasting especially during the periods of summer and fall.

Impacts of the frequency of cyanobacteria concertation sampling and other environmental variables on model efficiency were analyzed with different modeling time-steps. It was found that

15-day sampling frequency (or model simulation step) could be adequate for site-specific model development and reasonable short-term forecasting. Model 2 was validated for year 2018 at Cheney Reservoir with a 15-day data sampling interval and showed good model applicability. Model calibration with 15-day data sampling at Kansas River at Wamego showed good statistics for short-term forecasting with T, I-only Models 1 and 2.

Cyanobacteria growth rates were analyzed on the planes of environmental factors (T, I, P, N) and compared against observed data at both sites: Cheney Reservoir (in 2013-2014 and 2018) and Kansas River (in 2012-2014). Temperature and irradiation were found to strongly affect the growth rate. The effects of nutrients were relatively small due to the eutrophic condition of the studied water bodies. Evaluating growth rate patterns can help understand the seasonal variability of cyanobacteria in natural systems and respond to the changes in driving environmental factors.

Uncertainty in mechanistic model parameters and variability in observed cyanobacteria datasets were explored with a stochastic framework based on the assumption of influent nutrients affected by stream discharge fluctuations. The model was converted to a modified advection-dispersion equation in the form of the Fokker-Planck equation for cyanobacteria probability density function (PDF). A developed numerical solution with different initial log-normal distributions of PDF showed that uncertainty propagates to higher values of cyanobacteria concentrations with time. Higher uncertainty in initial cyanobacteria concentration led higher forecasted uncertainty. These findings can relate to the lower efficiency of the mechanistic model in long-term forecasting, compared to shorter forecasting periods and more frequent variable sampling.

5.2 Future Research

The implemented mechanistic and stochastic modeling approaches focused on describing cyanobacteria growth patterns at different temporal scales with limited available observed data in a large water body. The developed model considered cyanobacteria a single phytoplankton group while the impacts of other phytoplankton groups were ignored due absence of data. Expanding the framework to include different sampling locations within the lake, bacteria grazing patterns, and other phytoplankton groups can make the model more comprehensive and improve prediction of blooms within an entire lake. Moreover, coupling the cyanobacteria growth model with a lake algae transport model can describe spatial variability of blue green algae and better spot potential occurrences of CyanoHABs.

Different functional relationships were reported for cyanobacteria growth and T, I, P, and N variables. In our study, we considered the theta model and Monod kinetics model in equation (3.6). Other available functional relationships can be explored to examine their suitability for CyanoHAB dynamics modeling. It is worth noting that while some functional forms may be more efficient, they also require more data and input variables. The stochastic modeling approach can be extended in the future by incorporating random fluctuations of different physical variables to incorporate uncertainty embedded in measurements of temperature, irradiation, and nutrients.

References

Bowie, G. L., Mills, W. B., Porcella, D. B., Campbell, C. L., Pagenkopf, J. R., Rupp, G. L., ... Chamberlin, C.E. (1985). Rates, constants, and kinetics formulations in surface water quality modeling (2nd ed.). EPA, 600, 3-85.

Brient, L., Lengronne, M., Bertrand, E., Rolland, D., Sipel, A., Steinmann, D., . . . Bormans, M. (2008). A phycocyanin probe as a tool for monitoring cyanobacteria in freshwater bodies. *J. Environ. Monit., 10*(2), 248-255.

Canale, R. P., & Auer, M. T. (1982). Ecological studies and mathematical modeling of Cladophora in Lake Huron: 5. Model development and calibration. *J. Great Lakes Res., 8*(1), 112-125.

Canale, R. P., & Vogel, A. H. (1974). Effects of temperature on phytoplankton growth. *J. Environ. Eng. Div., 100*(1), 231-241.

Cerco, C. F., & Cole, T. (1994). Three-dimensional eutrophication model of Chesapeake Bay. *J. Environ. Eng., 119*(6), 1006-1025.

Chapra, S. C. (1997). Plant growth and nonpredatory Losses . In *Surface water-quality modeling*. (pp. 603–621). McGraw-Hill International editions.

Chapra, S. C., Boehlert, B., Fant, C., Bierman Jr, V. J., Henderson, J., Mills, D., ... Pearl, H. W. (2017). Climate change impacts on harmful algal blooms in US freshwaters: a screening-level assessment. *Environ. Sci. Technol., 51*(16), 8933-8943.

Cheney Lake Watershed Restoration and Protection Plan. (2011). Retrieved October 4, 2020, from http://www.kswraps.org/files/attachments/cheneylake_9e_plansummary.pdf

Chorus, I., & Bartram, J. (1999). Toxic cyanobacteria in water: A guide to public health significance, monitoring and management. World Health Organization. E&FN Spon, Routledge.–London, 416.

Chow-Fraser, P., Trew, D. O., Findlay, D., & Stainton, M. (1994). A test of hypotheses to explain the sigmoidal relationship between total phosphorus and chlorophyll a concentrations in Canadian Lakes. *Can. J. Fish. Aquat. Sci.*, *51*(9), 2052–2065.

Codd, G. A. (1995). Cyanobacterial toxins: Occurrence, properties and biological significance. *Water Sci. Technol., 32*(4), 149-156.

Dalu, T., & Wasserman, R. J. (2018). Cyanobacteria dynamics in a small tropical reservoir : Understanding spatio-temporal variability and influence of environmental variables. *Sci. Total Environ.*, 643, 835–841.

Dignum, M., Matthijs, H. C., Pel, R., Laanbroek, H. J., & Mur, L. R. (2005). Nutrient limitation of freshwater cyanobacteria. In *Harmful cyanobacteria* (pp. 65-86). Springer, Dordrecht.

Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., ... & Thornbrugh, D. J. (2009). Eutrophication of US freshwaters: Analysis of potential economic damages. *Environ. Sci. Technol.*, 43(1), 12-19.

Environmental Working Group (2020). News Reports of Algae Blooms, 2010 to Present. Retrieved September 9, 2020, from https://www.ewg.org/interactive-maps/algal_blooms/map/

Ferguson, A. J. D. (1997). The role of modelling in the control of toxic blue-green algae. *Hydrobiologia*, *349*(1-3), 1-4.

Fogg, G. E. (1969). The physiology of an algal nuisance. *Proc. of the Royal Society of London. Series B. Biological Sciences*, *173*(1031), 175-189.

Glibert, P. M., Allen, J. I., Bouwman, A. F., Brown, C. W., Flynn, K. J., Lewitus, A. J., & Madden, C. J. (2010). Modeling of HABs and eutrophication: Status, advances, challenges. *J. Mar. Syst.*, *83*(3-4), 262-275.

Graham, J.L., & Harris, T. D. (2016). Phytoplankton data for Cheney Reservoir near Cheney, Kansas, June 2001 through November 2015. Retrieved April 9, 2019, from https://www.sciencebase.gov/catalog/item/56fae421e4b0a6037df16848

Graham, J.L., Dubrovsky, N.M., & Eberts, S.M. (2017). Cyanobacterial harmful algal blooms and U.S. Geological Survey science capabilities (ver 1.1, December 2017): U.S. Geological Survey Open-File Report 2016–1174.

Graham, J.L., Ziegler, A.C., Loving, B.L., & Loftin, K.A. (2012). Fate and transport of cyanobacteria and associated toxins and taste-and-odor compounds from upstream reservoir releases in the Kansas River, Kansas, September and October 2011: U.S. Geological Survey Scientific Investigations Report 2012–5129.

Green Spanalga (Aphanizomenon flos-aquae). (2009). Retrieved September 10, 2020, from http://www.plingfactory.de/Science/Atlas/Kennkarten%20Procaryota/source/Aphanizomenon.html

Guven, B., & Howard, A. (2006a). Modelling the growth and movement of cyanobacteria in river systems. *Sci. Total Environ.*, *368*(2-3), 898-908.

Guven, B., & Howard, A. (2006b). A review and classification of the existing models of cyanobacteria. *Prog. Phys. Geogr.*, *30*(1), 1-24.

Harris, T. D., & Graham, J. L. (2016). Predicting cyanobacterial abundance , microcystin , and geosmin in a eutrophic drinking-water reservoir using a 14-year dataset. *Lake Reserv. Manag.*, *33(1)*, 32–48.

Harris, T. D., Smith, V. H., Graham, J. L., Van de Waal, D. B., Tedesco, L. P., & Clercin, N. (2016). Combined effects of nitrogen to phosphorus and nitrate to ammonia ratios on

cyanobacterial metabolite concentrations in eutrophic Midwestern USA reservoirs. *Inl. Waters*, 6(2), 199-210.

Havens, K. E., Ji, G., Beaver, J. R., Fulton, R. S., & Teacher, C. E. (2017). Dynamics of cyanobacteria blooms are linked to the hydrology of shallow Florida lakes and provide insight into possible impacts of climate change. *Hydrobiologia, 829*(1), 43-59.

Helsel, D. R., & Hirsch, R. M. (2002). *Statistical methods in water resources* (Vol. 49). Elsevier.

Ho, J. C., Michalak, A. M., & Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature, 574(7780),* 667-670.

Howard, A., Kirkby, M. J., Kneale, P. E., & McDonald, A. T. (1995). Modelling the growth of cyanobacteria (GrowSCUM). *Hydrol. Process.*, *9*(7), 809-820.

Huang, D. W., Wang, H. L., Feng, J. F., & Zhu, Z. W. (2008). Modelling algal densities in harmful algal blooms (HAB) with stochastic dynamics. *Appl. Math. Model.*, *32*(7), 1318-1326.

Hudnell, H. K. (Ed.). (2008). *Cyanobacterial harmful algal blooms: State of the science and research needs* (Vol. 619). Springer Science & Business Media.

Huisman, J., Matthijs, H. C. P., & Visser, P. M. (Ed.). (2005). *Harmful cyanobacteria* (Vol. 3). Springer, Dordrecht.

Janssen, A. B., Janse, J. H., Beusen, A. H., Chang, M., Harrison, J. A., Huttunen, I., ... Mooij, W. M. (2019). How to model algal blooms in any lake on earth. *Curr. Opin. Environ. Sustain.*, 36, 1-10.

Kansas Department of Health and Environment (2011). Allowances for low dissolved oxygen levels for aquatic life use. Water Quality Standards White Paper.

Kansas Department of Health and Environment (2020a). Harmful Algal Blooms: KDHE Agency Response Plan-2020. Retrieved from https://www.kdheks.gov/algae-illness/HAB_Response_Plan.htm

Kansas Department of Health and Environment (2020b). KDHE HABs Historical Data. Retrieved from https://www.kdheks.gov/algae-illness/historical_habs.htm

Kansas Department of Health and Environment (2020c). 2020 Kansas Integrated Water Quality Assessment. Retrieved from https://www.kdheks.gov/befs/download/Kansas_IR_2020_Final.pdf

Kong, Y., Lou, I., Zhang, Y., Lou, C. U., & Mok, K. M. (2014). Using an online phycocyanin fluorescence probe for rapid monitoring of cyanobacteria in Macau freshwater reservoir. *Hydrobiologia, 741*(1), 33-49.

Koning, Ross E. 1994. Cyanobacteria Slides. Plant Physiology Information Website. Retrieved September 10, 2020, from http://plantphys.info/organismal/lechtml/cyanobacteria.shtml

Malve, O., Laine, M., Haario, H., Kirkkala, T., & Sarvala, J. (2007). Bayesian modelling of algal mass occurrences-using adaptive MCMC methods with a lake water quality model. *Environ. Model. Softw.*, *22*(7), 966–977.

Monod, J. (1949). The growth of Bacterial Cultures. *Annu. Rev. Microbiol.*, *3*(Xl), 371–394.

New Hampshire Department of Environmental Services. (2017). Recreational Exposure to Cyanobacteria (Blue-Green Algae). Retrieved September 10, 2020, from https://www.des.nh.gov/organization/divisions/water/wmb/beaches/cyano_bacteria.htm

Paerl, H. W. (1988). Nuisance phytoplankton blooms in coastal, estuarine, and inland waters 1. *Limnol. Oceanogr.*, *33*(4part2), 823-843.

Paerl, H. W., & Fulton, R. S. (2006). Ecology of Harmful Cyanobacteria. In *Ecology of Harmful Algae* (pp. 95-109). Springer, Berlin, Heidelberg.

Paerl, H. W., & Huisman, J. (2009). Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. *Environ. Microbiol. Rep.*, *1*(1), 27–37.

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microb. Ecol.*, *65*(4), 995–1010.

Paerl, H. W., Hall, N. S., & Calandrino, E. S. (2011). Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Sci. Total Environ.*, *409*(10), 1739-1745.

Soboleva, T. K., & Pleasants, A. B. (2003). Population growth as a nonlinear stochastic process. *Math. Comput. Model.*, *38*(11-13), 1437-1442.

Protist Information Server. (n.d) Retrieved September 10, 2020, from http://protist.i.hosei.ac.jp/PDB/images/Prokaryotes/Chroococcaceae/Microcystis/sp_2c.jpg

Ralston, D. K., & Moore, S. K. (2020). Modeling harmful algal blooms in a changing climate. *Harmful Algae, 91,* 101729.

Reynolds, C. S., & Walsby, A. E. (1975). Water-blooms. *Biol. Rev.*, *50*(4), 437-481.

Reynolds, C. S., Oliver, R. L., & Walsby, A. E. (1987). Cyanobacterial dominance: the role of buoyancy regulation in dynamic lake environments. *New Zeal. J. Mar. Freshw. Res.*, *21*(3), 379-390.

Risken, H. (1989). *The fokker-planck equation: Methods of solution and application* (2nd ed.). Springer.

Scheffer, M., Rinaldi, S., Gragnani, A., Mur, L. R., & van Nes, E. H. (1997). On the dominance of filamentous cyanobacteria in shallow, turbid lakes. *Ecology*, *78*(1), 272-282.

Smith, V. H. (1985). Predictive models for the biomass of blue-green algae in lakes. *J. Am. Stat. Water Resour. Assoc., 21*(3), 433-439.

Smith, V. H., Sieber-Denlinger, J., de Noyelles, F., Campbell, S., Pan, S., Randtke, S. J., … Strasser, V. A. (2002). Managing taste and odor problems in a eutrophic drinking water reservoir. *Lake Reserv. Manag.*, *18*(4), 319–323.

Steele, J. H. (1965). Notes on some theoretical problems in production ecology. *Prim. Prod. Aqua. Environ*. Goldman, CR Berkley.

Stone, M.L., Graham, J.L., and Gatotho, J.W. (2013). Model documentation for relations between continuous real-time and discrete water-quality constituents in Cheney Reservoir near Cheney, Kansas, 2001–2009: U.S. Geological Survey Open-File Report 2013–1123.

United States Environmental Protection Agency (2017). National Water Quality Inventory: Report to Congress. Retrieved from https://www.epa.gov/waterdata/2017-national-water-quality-inventory-report-congress

Visser, P. M., Verspagen, J. M., Sandrini, G., Stal, L. J., Matthijs, H. C., Davis, T. W., ... & Huisman, J. (2016). How rising CO2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae*, *54*, 145-159.

Wang, P. (2011). Uncertainty quantification in environmental flow and transport models. PhD diss. California: The University of California, San Diego.

Wang, P., Tartakovsky, D. M., & Tartakovsky, A. M. (2013). Stochastic Forecasting of Algae Blooms in Lakes. In *Modelling and Simulation in Fluid Dynamics in Porous Media* (pp. 99-108). Springer, New York, NY.

Wang, S., Huggins, D. G., Lim, N. C., Baker, D. S., Spotts, W. W., Goodrich, C. A., ... Volkman, C.(2003). Cheney Reservoir water quality and its watershed assessment. Kansas Biological Survey.

Whitton, B. A. (Ed.). (2012). *Ecology of cyanobacteria II: Their diversity in space and time*. Springer Science & Business Media.

Whitton, B. A., & Potts, M. (Eds.). (2002). *The ecology of cyanobacteria: Their diversity in time and space*. Kluwer Academic Publishers.

Ziegler, A.C., Hansen, C.V., and Finn, D.A. (2010). Water quality in the *Equus* Beds aquifer and the Little Arkansas River before implementation of large-scale artificial recharge, south-central Kansas, 1995–2005: U.S. Geological Survey Scientific Investigations Report 2010–5023.

Zielinski, P. (1984). An application of the Fokker-Planck equation in stochastic reservoir theory. *Appl. Math. Comput., 15*(2), 123-136.

Appendix A Data Summary

Table A-1 Data for Cheney Reservoir

Date	Cyanobacteria (cells/ml)	Water Temperature (°C)	Irradiation (W/m^2)	Phosphorus (mg/L)	Nitrogen (mg/L)	Turbidity (NTU)	Dissolved Oxygen (mg/L)	Volume (acre-ft)
1/12/2013	872	1.68	116.13	0.038	0.948	5.14	14.12	102250.00
1/27/2013	809	3.84	118.57	0.042	0.932	6.63	13.13	101993.75
2/11/2013	988	3.87	137.92	0.048	0.993	8.33	12.89	102100.00
2/26/2013	968	3.71	167.33	0.057	1.032	11.11	13.21	103525.00
3/13/2013	1270	6.46	206.29	0.073	1.092	16.94	12.24	104575.00
3/28/2013	1460	8.49	186.62	0.092	1.176	24.69	11.32	104925.00
4/12/2013	1540	9.64	189.74	0.114	1.226	32.19	10.78	106762.50
4/27/2013	1200	12.39	251.22	0.104	1.087	21.56	9.95	109175.00
5/12/2013	610	18.74	255.15	0.105	1.013	17.69	8.64	111756.25
5/27/2013	1090	22.00	287.68	0.127	1.085	23.31	7.38	121387.50
6/11/2013	1080	24.31	303.78	0.152	1.038	30.58	6.49	125800.00
6/26/2013	3490	25.15	300.64	0.154	1.205	28.19	6.95	124618.75
7/11/2013	3770	25.75	247.16	0.152	1.205	25.88	6.59	123175.00
7/26/2013	5510	24.99	216.11	0.149	1.132	24.97	6.58	142325.00
8/10/2013	6320	24.83	264.33	0.125	0.903	17.25	5.99	209106.25
8/25/2013	1970	26.30	267.75	0.106	0.870	13.19	7.32	183075.00
9/9/2013	1800	24.29	200.80	0.122	0.966	20.69	7.02	167162.50
9/24/2013	2704	20.29	213.59	0.122	1.022	24.44	7.74	165062.50
10/9/2013	2730	16.19	171.03	0.111	1.072	23.94	8.91	162643.75
10/24/2013	2420	12.92	126.73	0.104	1.122	25.50	9.59	163800.00
11/8/2013	2160	9.50	116.64	0.096	1.118	25.88	10.43	165543.75
11/23/2013	1800	4.72	97.27	0.081	1.037	21.00	11.80	165275.00
12/8/2013	1540	0.91	86.23	0.066	0.949	15.56	13.04	165350.00
12/23/2013	1220	0.78	104.31	0.057	0.903	12.08	13.11	166568.75
1/7/2014	967	0.77	117.11	0.051	0.855	10.26	13.11	167350.00
1/22/2014	1012	0.52	106.64	0.056	0.907	12.43	13.67	167831.25
2/6/2014	1193	0.99	128.23	0.051	0.823	9.85	13.84	169481.25
2/21/2014	2470	1.29	151.50	0.052	0.873	9.24	13.79	171768.75
3/8/2014	3120	4.34	218.93	0.054	1.152	8.93	14.10	172000.00
3/23/2014	3570	8.24	206.20	0.059	1.124	9.16	11.57	170856.25
4/7/2014	1810	11.69	241.86	0.068	0.775	11.26	9.94	170350.00
4/22/2014	893	14.13	262.79	0.081	0.634	13.71	9.24	169050.00
5/7/2014	975	17.44	273.16	0.100	0.797	17.25	8.63	168343.75
5/22/2014	1130	21.14	266.74	0.085	0.819	9.95	7.52	168181.25
6/6/2014	2120	23.71	282.40	0.108	1.008	14.49	7.54	170425.00
6/21/2014	5420	25.13	271.28	0.115	1.005	14.94	6.64	173875.00
7/6/2014	4960	25.69	274.71	0.125	0.972	16.94	6.27	176556.25

Date	Cyanobacteria (cells/ml)	Water Temperature (°C)	Irradiation (W/m^2)	Phosphorus (mg/L)	Nitrogen (mg/L)	Turbidity (NTU)	Dissolved Oxygen (mg/L)	Volume (acre-ft)
7/21/2014	3680	25.82	283.72	0.119	0.813	14.75	6.29	172562.50
8/5/2014	7900	26.10	274.70	0.107	0.860	11.88	6.07	169737.50
8/20/2014	7680	25.68	236.62	0.112	1.033	14.50	6.34	165887.50
9/4/2014	6940	22.59	219.85	0.112	1.094	16.31	7.58	161675.00
9/19/2014	7270	21.01	199.87	0.097	1.040	13.69	7.62	158387.50
10/4/2014	7087	17.74	160.78	0.091	1.042	14.56	8.44	156343.75
10/19/2014	6710	15.97	152.71	0.081	1.021	13.13	8.80	155656.25
11/3/2014	6770	9.48	128.49	0.068	1.012	11.00	10.55	153350.00
11/18/2014	2930	4.45	106.35	0.053	1.035	7.51	12.01	152075.00
12/3/2014	1790	4.43	61.13	0.051	1.004	8.20	11.78	152675.00
12/18/2014	804	4.90	20.38	0.053	0.987	9.20	11.50	153200.00

Table A-2 Data for Kansas River (Wamego)

Date	Cyanobacteria (cells/ml)	Water Temperature (°C)	Irradiation (W/m^2)	Turbidity (NTU)	Dissolved Oxygen (mg/L)
8/31/2012	3935	24.58	224.28	7.90	9.35
9/15/2012	1432	19.93	180.91	7.90	9.43
9/30/2012	866	15.86	160.68	7.90	9.92
10/15/2012	378	14.51	149.53	7.90	10.18
10/30/2012	280	11.07	126.40	7.90	11.30
11/14/2012	262	8.17	116.64	7.90	10.11
11/29/2012	271	6.42	86.40	7.90	12.90
12/14/2012	239	2.21	85.67	7.90	13.84
12/29/2012	330	1.14	86.42	7.86	14.11
1/13/2013	399	2.05	116.79	4.91	13.93
1/28/2013	490	4.84	122.33	15.89	12.81
2/12/2013	415	2.85	130.94	14.23	13.29
2/27/2013	453	5.44	157.65	14.01	12.78
3/14/2013	298	7.33	177.84	16.94	12.57
3/29/2013	557	12.94	208.30	24.44	11.23
4/13/2013	1390	11.64	179.95	31.75	12.19
4/28/2013	1392	15.59	227.91	21.06	9.99
5/13/2013	934	21.04	264.59	17.81	8.88
5/28/2013	1026	21.43	280.91	23.69	7.80
6/12/2013	1220	24.88	263.11	30.75	7.34
6/27/2013	716	26.58	301.32	27.98	7.88
7/12/2013	1845	27.42	248.86	25.94	8.63
7/27/2013	2091	23.55	198.02	24.34	6.21

Date	Cyanobacteria (cells/ml)	Water Temperature (°C)	Irradiation (W/m^2)	Turbidity (NTU)	Dissolved Oxygen (mg/L)
8/11/2013	1234	25.78	260.99	17.06	7.52
8/26/2013	3092	28.32	260.81	13.19	8.41
9/10/2013	1312	23.31	200.32	21.50	8.74
9/25/2013	690	18.85	199.44	23.94	9.77
10/10/2013	670	14.43	155.39	24.00	10.67
10/25/2013	330	11.76	116.79	25.63	10.48
11/9/2013	237	7.83	108.87	25.75	11.23
11/24/2013	312	3.32	96.29	20.63	13.01
12/9/2013	535	1.63	84.39	15.00	13.36
12/24/2013	240	0.03	90.44	12.36	13.67
1/8/2014	232	0.48	106.23	10.44	12.48
1/23/2014	274	0.12	106.21	12.24	14.10
2/7/2014	244	0.84	122.98	9.83	13.11
2/22/2014	1365	2.29	154.82	9.96	13.31
3/9/2014	732	8.93	212.28	10.45	11.80
3/24/2014	985	9.91	187.11	9.67	11.57
4/8/2014	983	14.49	242.51	10.99	10.68
4/23/2014	3379	17.28	229.08	13.94	9.84
5/8/2014	3898	20.43	262.72	17.19	10.38
5/23/2014	4627	24.56	263.69	9.59	10.02
6/7/2014	4787	25.00	285.38	8.00	7.70

Appendix B Matlab Codes

B.1 Mechanistic model implementation

```
%Author- Md Atiqul Islam
% Code Objectives and scope: Optimize a HAB mechanistic model using
% historical data and compare model output with observations.
%% Reading collected historical data set from Xcel to Matlab:
clc
load Scaled_n=15.mat      %Contains 15 days averaged environmental
                          %variables and cyanobacteria concentrations @ each
                          %15th calendar day. Contains variable 'M' within
                          %this mat data file

%% Reading collected data set:
days=M(1:end,1);
DT1=flip((datetime(2013,1,12):caldays(1):datetime(2014,12,31))'); %Daily time
series
DT2=flip((datetime(2013,1,12):caldays(15):datetime(2014,12,31))'); %Bi-weekly
time series

%Reading 'M' from Scaled_n=15.mat
C=M(1:end,2);     %Cyanobacteria concentration (#/ml)
D=M(1:end,3);     %Outflow (ft^3/s)
Tu=M(1:end,4);     %Turbidity (NTU)
P=M(1:end,5);      %Total phosphorus (mg/L)
S=M(1:end,6);       %Suspended sediment (mg/L)
WT=M (1:end,7);     %water temperature (^oC)
E=M(1:end,8);       %Water elevation (ft)
N=M(1:end,9);        %Total nitrogen (mg/L)
I=M(1:end,10);        %GHI (W/m^2)
DO=M(1:end,11);        %Dissolvde oxygen (mg/L)
Chlorophyl=M(1:end,12);%Chlorophyll-a (microg/L)
Vol=M(1:end,13);         %reservoir volume (acre-ft)

%Assumed Parameters
sigma=0.27;
th_sigma=1.08;
alpha=9.52;
beta=10;
Tref=20;
AP=zeros(length(C),1); AP(1,1)=M(1,5)*1000-C(1)*alpha/1000;
AN=zeros(length(C),1);AN(1,1)=M(1,9)*1000-C(1)*beta/1000;

%%GR calculation
GR=zeros(length(C),4);
 for t=2:1:length(C)
    A=M(t,5)*1000-C(t)*alpha/1000; %converting to micro
    if A <= 0
        APP=0;
     else APP=A;
    end
    B=M(t,9)*1000-C(t)*beta/1000; %converting to micro
    if B <= 0
        ANN=0;
```

```
     else ANN=B;
     end
    Q=M(t,3)/(3.28)^3; %converting inflow to m^3/s
    V=M(t,13)*1233.48; %converting to m^3

    NGR=log(C(t-1)/C(t))/(days(t-1)-days(t));
    sigma_bar=sigma*(th_sigma^(WT(t)-Tref));
    wash=(Q*24*60*60/V);
    GR(t,1)=NGR+sigma_bar+wash;
    GR(t,2)=NGR;
    GR(t,3)=sigma_bar;
    GR(t,4)=wash;
AP(t)=APP;
AN(t)=ANN;
 end

 %%Data Fitting
 %x contains model inputs and b contains parameters
x=zeros(length(C),4);
x(:,1)=WT;
x(:,2)=I;
x(:,3)=P;
x(:,4)=C;
%x(:,5)=N; %for Model 4
y=GR(:,1);

fun = @(b,x)(b(1)*(b(2).^(x(:,1)-20)).*x(:,2)./(x(:,2)+b(3)).*(x(:,3)*1000-
x(:,4)*b(5)/1000)./((x(:,3)*1000-x(:,4)*b(5)/1000)+b(4))); %Model 3(TIP)
%fun = @(b,x)(b(1)*(b(2).^(x(:,1)-20)).*x(:,2)./(x(:,2)+b(3)).*(x(:,3)*1000-
x(:,5)*b(6)/1000)./((x(:,3)*1000-x(:,5)*b(6)/1000)+b(4)).*(x(:,4)*1000-
x(:,5)*b(7)/1000)./((x(:,4)*1000-x(:,5)*b(7)/1000)+b(5)));%Model 4(TIPN)

lb = [0.1,1.06, 1,1, 0.1];  %lower, upper and starting guess of mu, theta_mu,
Ki, Kp and alpha
ub = [0.50,1.16,400,60, 10];
x0 = [0.13,1.08, 100,5, 1];

[b,resnorm,residual,exitflag,output,lambda,J] = lsqcurvefit(fun,x0,x,y,lb,ub)

mdata = fun(b,x);   %model GR
sdata = GR(:,1);     %observation based GR
mmean = mean(sdata);

%Statistics based on GR
NSE1= 1- [(sum((sdata - mdata).^2))/(sum((sdata - mmean).^2))];
PBIAS1= sum(sdata - mdata)/sum(sdata)*100;
RSR1= [(sum((sdata - mdata).^2)).^0.5/(sum((sdata - mmean).^2)).^0.5];
RMSE1 = sqrt(mean((sdata - mdata).^2));

ci = nlparci(b,residual,'jacobian',J, 'alpha',0.95)   %Parametr 95% CI
[PreMean,delta] = nlpredci(fun,x,b,residual,'jacobian',J, 'alpha',0.95);
PreCI=zeros(length(PreMean),2);
PreCI(:,1) = PreMean - delta;     %GR prediction interval
PreCI(:,2) = PreMean + delta;
```

```
%Model Calibration plot
figure(1); clf
subplot(2,2,4)
scatter3(DT2,y,WT, [], WT,'o')
hold on
plot(DT2,fun(b,x),'-r')

ylabel('Growth Rate (per day)');
zlabel('Temperature(^0c)')
colorbar
legend('Observed','Model')

subplot(2,2,2)
scatter3(DT2,y,AN, [], AN,'o')
hold on
plot(DT2,fun(b,x),'-r')

ylabel('Growth Rate (per day)');
zlabel('Nitrogen(mg/L)')
colorbar

subplot(2,2,1)
scatter3(DT2,y,AP, [], AP,'o')
hold on
plot(DT2,fun(b,x),'-r')
colorbar

ylabel('Growth Rate (per day)');
zlabel('Phosphorus')

subplot(2,2,3)
scatter3(DT2,y,I, [], I,'o')
hold on
plot(DT2,fun(b,x),'-r')
colorbar

ylabel('Growth Rate (per day)');
zlabel('Irradiation')


%Patterrn Recognition on PN&WT plane
figure(2); clf
subplot(2,2,1)
scatter3(WT(2:end),AP(2:end),GR(2:end,1), [], AP(2:end)); hold on;
scatter3(WT, AP,PreMean, '*k');
xlabel('Temperature(^0c)')
ylabel('Available Phosphorus (mg/L)');
zlabel('Growth rate (per day)')
legend('Observed', 'Model')
colorbar
dim = [0.15 0.5 0.3 0.3];
str = {'Colorbar shows','Phosphorus(mg/L)'};
annotation('textbox',dim,'String',str,'FitBoxToText','on');

subplot(2,2,2)
```

```
scatter3(WT(2:end),AN(2:end),GR(2:end,1), [], AN(2:end)); hold on;
scatter3(WT, AN,PreMean, '*k');
xlabel('Temperature(^0c)')
ylabel('Nitrogen (mg/L)');
zlabel('Growth rate (per day)')
colorbar
legend('Observed', 'Model', '95% CI')
dim = [0.15 0.001 0.3 0.3];
str = {'Colorbar shows','Irradiation (W/m^2)'};
annotation('textbox',dim,'String',str,'FitBoxToText','on');

subplot(2,2,3)
scatter3(AP(2:end),GR(2:end,1), WT(2:end), [], WT(2:end)); hold on;
scatter3(AP,PreMean, WT,'*k');
xlabel('Available Phosphorus (mg/L)')
ylabel('Growth rate (per day)');
zlabel('Temperature(^0c)')
colorbar
legend('Observed', 'Model', '95% CI')
dim = [0.15 0.001 0.3 0.3];
str = {'Colorbar shows','Irradiation (W/m^2)'};
annotation('textbox',dim,'String',str,'FitBoxToText','on');

subplot(2,2,4)
scatter3(AN(2:end),GR(2:end,1), WT(2:end),[], WT(2:end)); hold on;
scatter3(AN,PreMean, WT,  '*k');
xlabel('Available Nitrogen (mg/L)')
ylabel('Growth rate (per day)');
zlabel('Temperature(^0c)')
colorbar
legend('Observed', 'Model', '95% CI')
dim = [0.15 0.001 0.3 0.3];
str = {'Colorbar shows','Irradiation (W/m^2)'};
annotation('textbox',dim,'String',str,'FitBoxToText','on');

%Patterrn Recognition on WT&I plane
figure(3); clf
subplot(3,1,1)
scatter3(WT(2:end),I(2:end),GR(2:end,1), [], I(2:end)); hold on;
scatter3(WT, I,PreMean, '*k');
xlabel('Temperature(^0c)')
ylabel('Irradiation (W/m^2)');
zlabel('Growth rate (per day)')
colorbar
dim = [0.15 0.001 0.3 0.3];
str = {'Colorbar shows','Irradiation (W/m^2)'};
annotation('textbox',dim,'String',str,'FitBoxToText','on');

subplot(3,1,2)
scatter3(I(2:end),GR(2:end,1),WT(2:end), [], WT(2:end)); hold on;
scatter3(I,PreMean, WT,'*k');
xlabel('Irradiation (W/m^2)')
ylabel('Growth rate (per day)');
zlabel('Temperature(^0c)')
colorbar
```

```
subplot(3,1,3)
scatter3(WT(2:end),GR(2:end,1),I(2:end), [], I(2:end)); hold on;
scatter3(WT,PreMean, I,'*k');
xlabel('Temperature(^0c)')
ylabel('Growth rate (per day)');
zlabel('Irradiation)')
legend('Observed', 'Model')
colorbar


%%BGA cell concentration Prediction

C_Pred=zeros(length(C),4);
C_Pred(:,1)= days;
C_Pred(length(C),2)= C(length(C),1);
C_Pred(length(C),3)= C(length(C),1);
C_Pred(length(C),4)= C(length(C),1);
for u=length(C)-1:-1:1
    AP=M(u+1,5)*1000-C(u+1)*alpha/1000; %converting to micro
    if AP <= 0
        AP=0;
     else AP=AP;
    end
    AN=M(u+1,9)*1000-C(u+1)*beta/1000; %converting to micro
    if AN <= 0
        AN=0;
     else AN=AN;
     end
    Q=M(u+1,3)/(3.28)^3; %converting to m^3/s
    V=M(u+1,13)*1233.48; %converting to m^3

    FNGR_Mean=PreMean(u+1)-(Q*24*60*60/V)-sigma*(th_sigma^(WT(u+1)-Tref));
    FNGR_LB=PreCI(u+1,1)-(Q*24*60*60/V)-sigma*(th_sigma^(WT(u+1)-Tref));
    FNGR_UB=PreCI(u+1,2)-(Q*24*60*60/V)-sigma*(th_sigma^(WT(u+1)-Tref));

    %For checking only
    Check_NGR(u+1,1)=PreMean(u+1);
    Check_NGR(u+1,2)=FNGR_Mean;
    Check_NGR(u+1,3)=sigma*(th_sigma^(WT(u+1)-Tref));
    Check_NGR(u+1,4)=(Q*24*60*60/V);
    Check_NGR(u+1,5)=FNGR_LB;
    Check_NGR(u+1,6)=FNGR_UB;

    %Long-term prediction (Without any update of cyanobacteria concentration)
    %C_Mean=C_Pred(u+1,2)*exp(FNGR_Mean*(days(u)-
days(u+1)));C_LB=C_Pred(u+1,3)*exp(FNGR_LB*(days(u)-
days(u+1)));C_UB=C_Pred(u+1,4)*exp(FNGR_UB*(days(u)-days(u+1)));

    %Short-term prediction (Update of cyanobacteria concentration @ regular
interval)
    C_Mean=C(u+1)*exp(FNGR_Mean*(days(u)-
days(u+1)));C_LB=C(u+1)*exp(FNGR_LB*(days(u)-
days(u+1)));C_UB=C(u+1)*exp(FNGR_UB*(days(u)-days(u+1)));

    C_Pred(u,2)= C_Mean;
```

```
    C_Pred(u,3)= C_LB;
    C_Pred(u,4)= C_UB;
                                end

%Cyanobacteria Prediction Confidence plot
figure(4); clf
plot(datenum(DT1(:,1)),bga_data(:,2),'b-', 'LineWidth', 1)
hold on

errorbar(datenum(DT2(1:end-1)),C_Pred(1:end-1,2),C_Pred(1:end-1,2)-
C_Pred(1:end-1,3),C_Pred(1:end-1,4)-C_Pred(1:end-1,2),'*k');
hold off
xlabel('Days')
ylabel('Cyanobacteria (#/ml)');
legend('Observed', 'Prediction Interval')
datetick('x','mmm-yy')

pdata = C_Pred(1:end-1,2); %Model Cyano
odata = C(1:end-1); %Observed Cyano
pdata_L = C_Pred(1:end-1,3);
pdata_U = C_Pred(1:end-1,4);

%1:1 Line predicted vs observed
figure(5); clf
errorbar(C(1:end-1),C_Pred(1:end-1,2),C_Pred(1:end-1,2)-C_Pred(1:end-
1,3),C_Pred(1:end-1,4)-C_Pred(1:end-1,2),'*k');
hold on
refline(1,0)
legend('95% prediction interval')
xlabel('Observed')
ylabel('Predicted');

%%Statistics Calculation based on Concentrations
mmean = mean(pdata);
NSE= 1- [(sum((pdata - odata).^2))/(sum((pdata - mmean).^2))]
PBIAS= sum(pdata - odata)/sum(pdata)*100
RSR= [(sum((pdata - odata).^2)).^0.5/(sum((pdata - mmean).^2)).^0.5]
RMSE = sqrt(mean((pdata - odata).^2))
```

B.2 Stochastic model implementation

```
%Author- Md Atiqul Islam
% Code Objectives and scope: Uncertainty quantification based on FP eqn.

clc
%% Problem formulations:
%Coeffiecients and Environmental variables
mu=0.334;
theta=1.0786;
T=21.13;                  %Temperature we are dealing
Tref=20;
sigma_bar=0.27*1.08^(T-Tref);
ki=37.78;
kp=2.37;
kn=21;
I=267;
alpha=4/1000;          %Divided by 1000 as mg/L to \micro gm/L conversion
issue
beta=10/1000;
Po=150;              %in \micro gm/L equivalent to mg/m^3
No=500;              %in \micro gm/L
Q=(2*66/3.28^3)*(24*60*60);      %Inflow volume in m^3 per day
q=(66/3.28^3)*(24*60*60);                %Outflow volume in m^3 per day
V=208*10^6;           %Lake volm in m^3

cp=733;                  %in mg/m^3
cn=2090;                  %in mg/m^3

%% Advection (D1) and Diffusion (D2) coefficients
%D1=(mu_bar - sigma_bar - q/V - p1*Cz)*C + [(dmu_bar^2/dQ*dC)*C +
(dmu_bar/dQ)]*[(dmu_bar/dQ)*C]
%D1=(T1 - sigma_bar - q/V - p1*Cz)*C + [T2*C + T3]*T3*C
%D2=[(dmu_bar/dQ)*C]^2=(T3*C)^2

%D1=0.3; D2=0.05;                            %Advection(v) & diffusion(k)
coefficients;

%% Non_Dimension
co=1130;
ND1=1/(mu*co);
ND2=V^2/(mu^2*co);
ND3=V^2/(mu^2*co^2);

%% Concentration and time boundary define
x_length=16000;          % Concentration limit
dx=10;                  % Step size for x
x=0:dx:x_length;
dx_st=dx/co;
x_st=0:dx_st:x_length/co;
%dt_stable= dx^2/(D1*dx+2*D2)

t_length=10;           % time duration
dt=1/100;                 % Step sixe for t
t=0:dt:t_length;
```

```
dt_st=dt*mu;
t_st=0:dt_st:t_length*mu;              %Dimensionless time

%Calculating partial derivatives
T1=zeros(length(t),length(x));
T2=zeros(length(t),length(x));
T3=zeros(length(t),length(x));

D1=zeros(length(t),length(x));              %Advection Coefficients
D2=zeros(length(t),length(x));              %Diffusion Coefficients

%% Implicit FD Solution for u(t,x)
%Initial Condition (IC)
w=zeros(length(t),length(x));
Co= 0;
C1=0;

w(1,:)=lognpdf(x_st,log(1),0.2);            %IC at time=0

%Coefficiients calculations
for j=2:1:length(t_st)
    P_bar(j-1,:)=Po+cp*(Q/V)*dt*((j-1)-1)-alpha*x;
    N_bar(j-1,:)=No+cn*(Q/V)*dt*((j-1)-1)-beta*x;
    T1(j-1,:)= mu*theta^(T-Tref)*(I/(ki+I))*(P_bar(j-1,:)/(kp+P_bar(j-
1,:)))*(N_bar(j-1,:)/(kn+N_bar(j-1,:)));
    T3(j-1,:)= (mu/V)*theta^(T-Tref)*(I/(ki+I))*[(P_bar(j-1,:)./(kp+P_bar(j-
1,:))).*(kn*cn./(kn+N_bar(j-1,:)).^2)+(N_bar(j-1,:)./(kn+N_bar(j-
1,:))).*(kp*cp./(kp+P_bar(j-1,:)).^2)];
    T2(j-1,:)= (mu/V)*theta^(T-Tref)*(I/(ki+I))*[(P_bar(j-1,:)./(kp+P_bar(j-
1,:))).*(2*beta*kn*cn./(kn+N_bar(j-1,:)).^3)+(N_bar(j-1,:)./(kn+N_bar(j-
1,:))).*(2*alpha*kp*cp./(kp+P_bar(j-1,:)).^3)-
(kp*kn*(beta*cp+alpha*cn)./((kp+P_bar(j-1,:)).^2.*(kn+N_bar(j-1,:)).^2)) ];
    D1(j-1,:)= ND1*((T1(j-1,:) - sigma_bar - q/V - p1*Cz).*x) + ND2*([T2(j-
1,:).*x + T3(j-1,:)].*T3(j-1,:).*x);
    D2(j-1,:)=ND3*((T3(j-1,:).*x).^2);

end
    P_bar(length(t),:)=Po+cp*(Q/V)*dt*((length(t))-1)-alpha*x;
    N_bar(length(t),:)=No+cn*(Q/V)*dt*((length(t))-1)-beta*x;
    T1(length(t),:)= mu*theta^(T-
Tref)*(I/(ki+I))*(P_bar(length(t),:)/(kp+P_bar(length(t),:)))*(N_bar(length(t
),:)/(kn+N_bar(length(t),:)));
    T3(length(t),:)= (mu/V)*theta^(T-
Tref)*(I/(ki+I))*[(P_bar(length(t),:)./(kp+P_bar(length(t),:))).*(kn*cn./(kn+
N_bar(length(t),:)).^2)+(N_bar(length(t),:)./(kn+N_bar(length(t),:))).*(kp*cp
./(kp+P_bar(length(t),:)).^2)];
    T2(length(t),:)= (mu/V)*theta^(T-
Tref)*(I/(ki+I))*[(P_bar(length(t),:)./(kp+P_bar(length(t),:))).*(2*beta*kn*c
n./(kn+N_bar(length(t),:)).^3)+(N_bar(length(t),:)./(kn+N_bar(length(t),:))).
*(2*alpha*kp*cp./(kp+P_bar(length(t),:)).^3)-
(kp*kn*(beta*cp+alpha*cn)./((kp+P_bar(length(t),:)).^2.*(kn+N_bar(length(t),:
)).^2)) ];
    D1(length(t),:)= ND1*((T1(length(t),:) - sigma_bar - q/V - p1*Cz).*x) +
ND2*([T2(length(t),:).*x + T3(length(t),:)].*T3(length(t),:).*x);
    D2(length(t),:)=ND3*((T3(length(t),:).*x).^2);
```

```
%Co-efficients a, b, c and f
a=zeros(length(t),length(x));
b= zeros(length(t),length(x));
c= zeros(length(t),length(x));

for j=1:1:length(t_st)
    for m=1:1:length(x_st)
a(j,m)= -(D1(j,m)/dx_st) - D2(j,m)/dx_st^2;
b(j,m)= (1/dt_st)+(D1(j,m)/dx_st) + 2*D2(j,m)/dx_st^2;
c(j,m)= -D2(j,m)/dx_st^2;
    end
end


% Matrix solver
A=zeros(length(x),length(x));
A(1,1)=1/dt_st;                     %Applying BC in matrix A
A(end,end)=1/dt_st;                 %Applying BC

M=zeros(length(t),3);       %Contains maximum PDF & concentration @specific
time
[maxpdf, con_ID] = max(w(1,:));
    M(1,1)=t(1);                  %time
    M(1,2)=x_st(:,con_ID);                  %concentration
    M(1,3)=maxpdf;                  %maxm pdf

figure(3);
for j=1:1:length(t)-1
   for m=2:1:length(x)
       A(m,m-1)=a(j+1,m);
       A(m,m)=b(j+1,m);
       A(m,m+1)=c(j+1,m);
   end
    A(1,1)=b(j+1,1);
    A(1,2)=c(j+1,1);
    f= transpose (w(j,:))./(dt_st);
    solution = A(:, 1:end-1)\f;
    %solution = pinv(A(:, 1:end-1))*f;
    w(j+1,:) = transpose (solution);
    %w(j+1,1) =C1;
    %w(j+1,end) =Co;
    [maxpdf, con_ID] = max(w(j+1,:));
    M(j+1,1)=t(j+1);                  %time
    M(j+1,2)=x_st(:,con_ID);                  %concentration
    M(j+1,3)=maxpdf;                  %maxm pdf
end

wsol=w;
 plot(x_st, w(1:100:end,:), '-b');


figure (1); clf
plot(x_st, w(1,:),'b-', 'LineWidth',1.5); %Initial condition plot at time=0
```

```
hold on;
plot(x_st, w(501,:),'r--','LineWidth',1.5); %PDF @ time=5 days
hold on;
plot(x_st, w(end,:),'g-.','LineWidth',1.5); %PDF @ time=10 days
hold on;


        xlabel('Dimensionless Concentration');
        ylabel('PDF (W)');

 legend('t=0 day','t=5 days','t=10 days');
 title('Advection Diffusion of PDF in Time-Conccentration with lognormal
distribution lognpdf(x_st,log(1),0.2) and co=1130 cells/ml');

figure (2); clf
plot(x, w(1,:),'b-', 'LineWidth',1.5); %Initial condition plot at time=0
hold on;
%plot(x_st, w(100:100:end,:), '-g');
plot(x, w(501,:),'r--', 'LineWidth',1.5); %PDF @ time=5 days
hold on;
plot(x, w(end,:),'g-.', 'LineWidth',1.5); %PDF @ time=10 days
hold on;


        xlabel('Cyanobacteria Concentration (cells/ml)');
        ylabel('PDF (W)');

 legend('t=0 day','t=5 days','t=10 days');
 title('Advection Diffusion of PDF in Time-Conccentration with lognormal
distribution lognpdf(x_st,log(1),0.2) and co=1130 cells/ml');

figure(4); clf
[hAx,hLine1,hLine2] = plotyy(M(:,1), M(:,2),M(:,1), M(:,3));
title('Maximum PDF, Cyanobacterial Growth vs Time')
xlabel('Time (days)')
ylabel(hAx(1),'Concentration Dimensionless') % left y-axis
ylabel(hAx(2),'PDF') % right y-axis
legend('Concentration', 'PDF')

save(sprintf('PDF_Q=%g.mat', 2),'w'); %saving output file
```

Appendix C Derivatives in the Fokker-Planck Equation

The following relationships were used in equations (4.9a) - (4.9b). Values used for various parameters and environmental variables presented in Figure 4.1-4.2 are listed in Table C-1.

$$GR = \mu_m \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I} \frac{\bar{P}}{K_P + \bar{P}} \frac{\bar{N}}{K_N + \bar{N}} \tag{C.1}$$

$$NL = \sigma_m \theta_\sigma^{T-T_{Ref}}, \qquad WO = \frac{q_{out}}{V} \tag{C.2}$$

$$\frac{\partial GR}{\partial Q} = \frac{\mu_m}{V} \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I} \left[\frac{K_P c_P}{(K_P + \bar{P})^2} \frac{\bar{N}}{K_N + \bar{N}} + \frac{\bar{P}}{K_P + \bar{P}} \frac{K_N c_N}{(K_N + \bar{N})^2} \right] \tag{C.3}$$

$$\frac{\partial^2 GR}{\partial Q \partial \xi} = \frac{\mu_m}{V} \theta_\mu^{T-T_{Ref}} \frac{I}{K_I + I} \left[\frac{2\alpha K_P c_P}{(K_P + \bar{P})^3} \frac{\bar{N}}{K_N + \bar{N}} + \frac{\bar{P}}{K_P + \bar{P}} \frac{2\beta K_N c_N}{(K_N + \bar{N})^3} - \frac{K_P K_N (\beta c_P + \alpha c_N)}{(K_P + \bar{P})^2 (K_N + \bar{N})^2} \right] \tag{C.4}$$

$$\bar{P} = P_0 + c_P \frac{Q}{V} - \alpha\xi, \qquad \bar{N} = N_0 + c_N \frac{Q}{V} - \beta\xi \tag{C.5}$$

Table C-1 Definitions of parameters and variables

	Definition	Values	Unit
Parameters			
μ_m	maximum growth rate at 20℃	0.334	day^{-1}
σ_m	maximum non-predatory loss rate at 20℃	0.27	day^{-1}
θ_μ	temperature coefficients for growth rate	1.0786	-
θ_σ	temperature coefficients for non-predatory loss rate	1.08	-
K_I	global irradiance half-saturation coefficient	37.78	$W\ m^{-2}$
K_P	phosphorus half-saturation coefficient	2.37	$mg\ m^{-3}$
K_N	nitrogen half-saturation coefficient	21	$mg\ m^{-3}$
α	relative phosphorus content of cyanobacteria	0.04	-
β	relative nitrogen content of cyanobacteria	0.12	-
Variables			
ξ	random cell concentration of cyanobacteria	1130	$cells\ L^{-1}$
$\bar{P}$	average available phosphorus for cyanobacteria	-	$mg\ m^{-3}$
P_0	initial phosphorus concentration	150	$mg\ m^{-3}$
c_P	inflow phosphorus concentration	733	$mg\ m^{-3}$
$\bar{N}$	average available nitrogen for cyanobacteria	-	$mg\ m^{-3}$
N_0	initial nitrogen concentration	500	$mg\ m^{-3}$
c_N	inflow nitrogen concentration	2090	$mg\ m^{-3}$
T, T_{ref}	temperature, the reference temperature (20℃)	21.13, 20	℃
I	global horizontal irradiance	267	$W\ m^{-2}$
Q	Inflow volume	323,196	$m^3 day^{-1}$
q_{out}	Outflow volume	161,598	$m^3 day^{-1}$
V	volume of lake	208,000,000	m^3