Assessment of Drivers of Algal Biomass in North American Great Lakes via Satellite Remote Sensing

by

Michael Arthur Dallosch

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Geography

Waterloo, Ontario, Canada, 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner   Dr. Ben DeVries
                                     University of Guelph – Department of
                                     Geography, Environment and Geomatics

Supervisor(s)   Dr. Claude Duguay
                                     University of Waterloo – Department of
                                     Geography and Environmental Management

                                     Dr. Homa Kheyrollah Pour
                                     Wilfrid Laurier University - Department of
                                     Geography and Environmental Studies

Internal Member   Dr. Merrin Macrae
                                     University of Waterloo – Department of
                                     Geography and Environmental Management

Internal-External Member   Dr. Roland Hall
                                     University of Waterloo – Department of
                                     Biology

Other Member(s)   Dr. Kiana Zolfaghari
                                     University of Waterloo – Department of
                                     Geography and Environmental Management

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Lakes are regarded as sentinels of change, where shifts in environmental conditions significantly affect lake phenology. A significant consequence of the change is the perceived increase in the frequency, magnitude, and severity of algal blooms in lakes globally. Algal blooms/increased productivity in lakes pose significant ecological, economic and health risks, impacting fisheries, tourism, and freshwater access. The impacts of external nutrient loading from anthropogenic sources are well documented; however, blooms have been observed to occur in even remote lakes. Climate change is a hypothesized driver of these recent algal bloom trends, such as increasing global air temperatures, water temperatures, lake ice loss, precipitation intensity, and drought. Past research on the impact of climatic drivers on algal biomass dynamics has often been limited to lab, mesocosm, or short termed observations, due to limited in situ data. New remote sensing data products make use of historic multispectral satellite image archives to provide greater spatial and temporal coverage of algal biomass concentrations, allowing for longer time series observational studies to be conducted over large areas. Using data provided by the European Space Agency (ESA) Climate Change Initiative (CCI) Lakes project (product version 2.0.0), daily chlorophyll-*a* (chl-*a*; proxy of algal biomass), Lake Surface Water Temperature (LSWT) and Lake Ice Cover (LIC) from 2002 to 2020 were derived from five North American Great Lakes: Great Bear Lake (GBL), Great Slave Lake (GSL), Lake Athabasca (LA), Lake Winnipeg (LW), and Lake Erie (LE). Additional atmospheric and lake physical variables were provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-Land data as part of the ERA5 climate reanalysis product including: 2-m air temperature (T2m), Total Precipitation (PPT), Surface Net Solar Radiation (SNSR), Surface Runoff (SR) and Subsurface Runoff (SSR), Wind Speed (WS) and Lake Mix-Layer Depth (LMLD). Such data products allow for comprehensive time series analysis on the interaction effects of atmospheric and lake physical parameters on algal biomass dynamics.

Winter temperatures exhibit the highest rate of change relative to other seasons, where LIC loss is important for Northern hemisphere lakes; however, its effect on algal biomass

dynamics is relatively unknown. To investigate how LIC duration alters algal biomass in North American Great Lakes, annual and seasonal algal biomass, LSWT and LIC parameters were calculated for the five study lakes using ESA CCI Lakes data. Algal biomasses (β = 0.01 – 0.75 µg L^{-1} yr^{-1}) and LSWT (β = 0.03 – 0.14 K yr^{-1}) were found to increase, with a general decrease in LIC (β = -0.88 – -1.08 Days yr^{-1}) from 2002 to 2020. Vector autoregressions (VARs) showed that in Northern Lakes (NL; GBL, GSL and LA), LSWT and LIC parameters provide greater explanatory power for annual/seasonal chl-a concentrations ($\widetilde{adj.\ r^2}$ = 0.75) compared to Southern Lakes ($\widetilde{adj.\ r^2}$ = 0.46). Additionally, LIC parameters were found to provide higher explanatory power for NLs during the spring season compared to LSWT. However, higher explanatory power does not indicate predictive capacity, where machine learning methods may provide stronger predictive models.

To determine if LIC may act as a predictor of algal biomass parameters, multiple linear regression (MLR) and artificial neural networks (ANN) were constructed using per-pixel observations of annual/seasonal algal biomass, LSWT, and LIC parameters. Irrespective of season, LSWT only models returned lower prediction error (median NRMSE = 0.82) compared to LIC only models (median NRMSE = 0.93). However, models consisting of both LIC and LSWT returned the lowest predictive error (median NRMSE = 0.75). While LIC did not act as a strong predictor of algal biomass, a random forest (RF) classifier was used to determine whether LIC could classify the presence of lake-specific anomalies in chl-a concentrations. The RF model found that LIC parameters (ice on/off) had the highest mean accuracy decrease on average for NLs during the spring season. LIC timings are changing, where it was found to have greater importance on springtime abnormal algal biomass growth in NLs. While LIC was important at this time compared to LSWT, the impact of other important atmospheric and lake physical variables on algal biomass dynamics are not well understood, particularly at a smaller temporal scale (*i.e.*, daily).

To assess the potential interaction effects between algal biomass, atmospheric, and lake physical parameters, a network analysis was conducted using a High Order Dynamic Gaussian Bayesian Network (HO-DGBN) for the original time series, the stationary, non-stationary, and

residual signals at varying temporal ranges (Δ: daily, three days, weekly, biweekly, and monthly averages). It was found that LSWT, T2m and SNSR were the most important parameters on average, where LSWT exhibited the highest importance on the daily scale compared to the monthly. Additionally, LMLD returned increased importance at longer temporal frequencies, while SSR returned increased importance at shorter temporal frequencies. Temperature interactions were mixed, typically returning both positive and negative interactions, while SNSR typically exhibited a positive interaction with chl-*a*, while LMLD exhibited a frequent negative interaction. PPT and WS were found to be the least important parameters in all study lakes.

This thesis provides some of the first analytical uses of the ESA CCI Lakes product; a product that undergoes regular updates (every two years or so) as new satellite and *in situ* data become available, and algorithms for the retrieval of chl-*a*, LSWT and LIC are being improved. As such, improvements are expected in future releases of the product, limiting the accuracy of some findings in the thesis. Of the data presented, there is evidence that LIC is a significant contributor to spring algal biomass dynamics for NLs; however, Southern Lakes (SL; LW and LE) exhibit more complex interactions, likely due to anthropogenic impacts. This thesis identifies the complexity of LSWT interactions with algal biomass and identifies LMLD as a predominantly negative effect in the development of algal biomass. Algal biomasses are increasing, where increases in LSWT yield higher algal biomass peaks (at varying times throughout the year) within the study lakes. Future climate scenarios may provide conditions favorable for algal biomass growth, where Northern landscapes are at the greatest risk.

Acknowledgements

Table of Contents

List of Figures

List of Tables

xxv

List of Abbreviations

6S	Second Simulation of a Satellite Signal in the Solar Spectrum
AB	All Basins
ABs	Algal Blooms
Adj. r²	Corrected (adjusted) coefficient of determination
AICc	Akaike information criterion second order
ANN	Artificial Neural Network
AOP	Apparent Optical Properties
AT	Anomaly Lake Surface Water Temperature
BC	Abnormal Chlorophyll-*a* Concentrations
BD	Abnormal Chlorophyll-*a* Concentrations Days
BIC	Bayesian Information Criteria
BN	Bayesian Network
C	Carbon
CB	Central Basin
CDOM	Coloured Dissolved Organic Matter
CH$_4$	Methane
Chl-*a*	Chlorophyll-*a*
CO$_2$	Carbon Dioxide
Cu	Copper
DAG	Directed Acyclic Graph
DDV	Dark Dense Vegetation
DEM	Digital Elevation Model
DGBN	Dynamic Gaussian Bayesian Network
DMMHC	Dynamic Max-Min Hill-Climbing
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter
DOS	Dark Object Subtraction
DOY	Day of Year
DR	Ice Duration
EB	East Basin
ECMWF	European Centre for Medium-Range Weather Forecasts

ESA CCI	European Space Agency Climate Change Initiative
ETM+	Enhanced Thematic Mapper Plus
Fe	Iron
Fe^{2+}	Ferrous Iron
Fe^{3+}	Ferric Iron
FPE	Forecast Prediction Error
GBL	Great Bear Lake
GBN	Gaussian Bayesian Network
GS	Grow-Shrink
GS	Greedy Search
GSL	Great Slave Lake
H_2CO_3	Carbonic Acid
HC	Hill-Climbing
HCO_3-	Bicarbonate
HO-DGBN	High-Order Dynamic Gaussian Bayesian Network
IAMB	Incremental Association
IOP	Inherent Optical Properties
LA	Lake Athabasca
LASSO	Least Absolute Shrinkage and Selection Operator
LE	Lake Erie
LIC	Lake Ice Cover
LMLD	Lake Mixing Level Depth
LW	Lake Winnipeg
LWLR	Lake Water Leaving Reflectance
LWST	Lake Surface Water Temperature
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Median Absolute Percentage Error
MC	Mean Chlorophyll-*a* Concentrations
MDA	Mean Decrease Accuracy
MERIS	MEdium Resolution Imaging Spectrometer
MK	Mann-Kendall
MLR	Multiple Linear Regression
MMHC	Max-Min Hill-Climbing
MODIS	MODerate resolution Imaging Spectroradiometer

MT	Mean Lake Surface Water Temperature
N	Nitrogen
N₂O	Nitrous Oxide
NAP	Non-algal Particles
NB	North Basin
NIR	Near Infrared
NRMSE	Normalized Root Mean Squared Error
OAC	Optically Active Constituent
OF	Ice Off Day of Year
OLI	Operational Land Imager
ON	Ice On Day of Year
OWT	Optical Water Type
P	Phosphorus
PC	Peak Chlorophyll-*a* Concentrations DOY
PSO	Particle Swarm Optimization
PT	Peak Lake Surface Water Temperature DOY
RF	Random Forest
RJ-MCMC	Reversible Jump Markov Chain Monte Carlo
RMSE	Root Mean Squared Error
SB	South Basin
SELU	Scaled Exponential Linear Units
Si	Silicon
SIMoNe	Statistical Inference for Modular Networks
SLFN	Single Layer Feedforward Neural Network
SNR	Signal-to-Noise ratio
SNSR	Surface Net Solar Radiation
SR	Surface Runoff
SSR	Subsurface Runoff
T2m	2m Air Temperature
TM	Thematic Mapper
TP	Total Precipitation
TS	Tabu Search
T-S	Theil-Sen
VAR	Vector Autoregression
WB	West Basin

WS	Wind Speed
XC	Max Chlorophyll-*a* Concentrations
XT	Max Lake Surface Water Temperature
Zn	Zinc

List of Symbols

$\mathbf{L}_\lambda$	Top-of-atmosphere Radiance
$\mathbf{E}_{sun}$	Exo-atmospheric Irradiance
$t_{oz}\downarrow$	Downwelling ozone transmittance
$t_{oz}\uparrow$	Upwelling ozone transmittance
λ	Wavelength
θ_s	Solar zenith angle
θ_v	Satellite viewing angle
ϕ_s	Solar elevation angle
ϕ_v	Satellite elevation angle
ψ_s	Solar azimuth angle
ψ_v	Satellite azimuth angle
$R(0^-,\lambda)$	Subsurface spectral irradiance
$a_i(\lambda)$	absorption coefficient at a given wavelength
b_b	backscatter coefficient
$\mathcal{L}$	Random forest dataset
$\mathcal{L}_k$	Bagged data
L	Total hidden nodes
$g_i(x_j)$	Activation function
b_i	Bias of each hidden node
a_{iN}	Weights to the ith hidden node

X_i	Given variable
Π_{X_i}	Parents of X_i
x_j	Target node
Pa_j	Set of parent nodes for x_j
$\mathcal{N}$	Density function of a normal distribution
β_{0j}	Independent coefficient
σ_j^2	Variance of x_j
$r(i)$	Total amount of parent nodes for x_j
t	Time-slice
X_t	Vector of all the nodes in a time slice t
T	Total time slices
β	Slope | Beta for Holts-Winter forecasting
s_e	Standard error
p	P-value
r^2	Coefficient of determination
y_t	Set of K endogenous variables
A_i	Coefficient $(K \times K)$
u_t	K dimensional vector of the t residuals
D_t	Deterministic regressors
$\mu g\ L^{-1}$	Micrograms per Liter
K	Kelvin

α	Multiple linear regression bias | alpha for Holts-Winter forecasting
$\beta_{1\cdots n}$	Multiple linear regression weights
μ	Mean
σ	Standard deviation
O	Outliers
$Q3$	Third quartile
γ	Gamma
$\widehat{Chla}_{reconstructed}$	Reconstructed time series of chl-*a* concentrations
$\widehat{Chla}_{stationary}$	Stationary time series of chl-*a* concentrations
$\widehat{Chla}_{nonstationary}$	Non-stationary time series of chl-*a* concentrations
$\widehat{Chla}_{residuals}$	Residuals of chl-*a* concentration time series
Δ	Temporal frequency

Chapter 1: Introduction

1.1 Motivation

There is a perception that the frequency, magnitude and severity of algal blooms have increased globally in recent decades, posing significant environmental, economic and human health issues (Paerl & Huisman, 2008; Winter *et al.,* 2011; Brooks *et al.,* 2016; Pick, 2016; Favot *et al.,* 2023). There has been a larger increase in blooms in recent years compared to those of past decades (Hou *et al.,* 2022). Phytoplankton are ecologically important for lakes, as they form the basis of aquatic food webs (Field *et al.,* 1998; Wilken *et al.,* 2018), are significant contributors of global biogeochemical cycling (Carr & Whitton, 1982; Wehr *et al.,* 2015), and contribute ~40% of atmospheric oxygenation (Falkowski and Raven, 2007). Additionally, phytoplankton are photosynthetic autotrophs, and key in the storage and cycling of phosphorus, cycling of iron, and the fixation/remineralization of nitrogen (Gimenez *et al.,* 2018). Lakes are considered sentinels of change, where the shift in algal species composition and biomass is key to indicators of biogeochemical and lake physical changes (Adrian *et al.,* 2009; Kong *et al.,* 2017a). The excess of algae leads to increased bacterial respiration of DO during algal decomposition, resulting in deep-water anoxia (Wells *et al.,* 2015). Shifts in algal community structure, introduction of invasive species, and anoxia may also limit or suppress native zooplankton (Roman *et al.,* 2019; Senar *et al.,* 2019). Blooms may result in beach and lake closures, impacting tourism and water sports industries, while anoxia can lead to fish kills, significantly impacting fisheries (Moore *et al.,* 2019; Smith *et al.,* 2019). Nuisance blooms can clog water intake pipes and disrupt drinking water availability, while many genera of cyanobacteria produce neurotoxins which can result in neurological, hematological, and hepatological illnesses (Hilborn & Beasley, 2015).

The trend in increasing algal biomass and bloom formation is traditionally attributed to the anthropogenic introduction of allochthonous macro and micronutrients (Schindler, 1977; Huisman *et al.,* 2005). However, in recent years, blooms have been observed to occur within remote and oligotrophic lakes with minimal anthropogenic influence (Pick, 2016; Favot *et al.,* 2019; Smol, 2019). Global climate change is an anticipated driver of these recent trends in

algal biomass (Paerl & Huisman, 2008; Bartosiewicz *et al.,* 2019; Ho & Michalak, 2020). There has been an increase in global air temperatures of ~1.09°C (Pörtner *et al.,* 2022), with high-latitude regions experiencing the greatest increase (Screen & Simmonds 2010; Jeong et al., 2014). Higher air temperatures drive freshwater warming, where many genera of algae yield a higher growth rate (Wells *et al.,* 2015). Changes in temperature can also affect the rate and duration of stratification (Kraemer *et al.,* 2015), lake mixing regimes (Adrian *et al.,* 2009; Ficker *et al.,* 2017), and even nutrient exchange, which can affect algae in different ways (Tammeorg *et al.,* 2020; Harrow-Lyle & Kirkwood, 2021). Shifts in precipitation affect the rate of external loading (Creed & Band 1998; Coffey *et al.,* 2019), flushing (Ho & Michalak, 2020) and mixing (Bakker & Hilt, 2016) within a lake, which ultimately affects algal biomass growth (Reichwaldt & Ghadouani, 2012). Factors such as wind speed and light availability will also impact algal growth dynamics (Huang *et al*., 2016; Zhou *et al*., 2021). The presence and duration of lake ice cover is a key factor in controlling lake mixing regimes, light availability, and the duration of the growing season (Rumyantseva *et al.,* 2019). Canadian winter temperatures are increasing at a higher rate than in any other season (Zhang *et al.,* 2019a), thus disproportionately impacting overwintering conditions. As such, it is imperative to understand how this change in ice cover may impact algal biomass.

There have been many studies that have addressed the impacts of atmospheric and lake physical variables on algal biomass dynamics; however, these are often limited to lab-controlled experiments, mesocosm studies, and short-term observations, due to data limitations. In addition, very few observational studies have been conducted on the connection between changes in lake ice cover and algal biomass. The effects of lake ice cover change on algal biomass are relatively unknown. New remote sensing data products and climate reanalysis data sets have allowed comprehensive observational studies to better understand the dynamics between climate-induced variables and algal biomass. Such research may be used to improve algal biomass forecasts, projection models, and early warning systems.

1.2 Objectives

The primary objective of this research is to determine the influence that climate-induced atmospheric and lake physical variables have on the trends, variability, and dynamics of algal biomass in five North American Great Lakes: Great Bear Lake (hereafter referred to as GBL), Great Slave Lake (GSL), Lake Athabasca (LA), Lake Winnipeg (LW) and Lake Erie (LE) over a 19-year period (2002 – 2020). These lakes are selected to provide varying latitudinal and anthropogenic conditions, where the Northern Lakes (NL; GBL, GSL, and LA) reside within primarily natural watersheds with minimal anthropogenic development, while the Southern Lakes (SL; LW and LE) reside within heavily developed watersheds (agriculture and urban). Due to limited *in situ* observations, complete time series of lake biological and physical characteristics (*e.g*., algal biomass, lake ice, surface water temperatures, etc.) across many lakes (particularly in Northern regions) are not available. Utilizing new remote sensing data products provided by the European Space Agencies (ESA) Climate Change Initiatives (CCI) Lakes project, it is possible to provide a comprehensive analysis of algal biomass trends, variabilities, and dynamics. The ESA CCI Lakes product provides a harmonized 1-km daily gridded imagery from 2002 – 2020 of various lake parameters including lake ice classification, chlorophyll-*a* concentration (chl-*a*; proxy of algal biomass) and lake surface water temperatures (LSWT). Additional gridded climatic parameters such as precipitation, wind speed, net surface solar radiation, land air temperature, and surface/sub-surface runoff provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-land allow for additional interaction effects analysis.

This research aimed to achieve three main objectives: (1) determine if lake ice affects the trends and variability of algal biomass, (2) determine if lake ice is a predictor of algal biomass, and (3) assess and compare the interaction effects of lake physical and atmospheric forcings on algal biomass dynamics.

1.3 Structure

This thesis is comprised of a total of seven chapters and structured in a manuscript format. This first and current chapter highlights the general motivations, purpose, and

objectives regarding algal biomass dynamics, its environmental drivers, and how remote sensing products may be used to fill in current data gaps. Chapter 2 provides an in-depth review of how environmental variables associated with climate change affect the abiotic and biotic drivers within a lake, thus affecting the growth rate, concentration, and community composition of algae in freshwater lakes. Chapter 2 also provides information regarding how these variables are changing and their implications for lake phenology, and a review of remote sensing techniques and challenges within freshwater systems, providing an overview of empirical, machine learning, and bio-optical chl-*a* retrieval techniques. Additional information is provided on Bayesian network analysis techniques to provide additional context for the methodology of Chapter 6. Chapter 3 covers an overview of the study area and data sets used for this thesis. All research conducted in the three following chapters (4 – 6) was carried out for the same five North American Great Lakes and used the same ESA CCI Lakes and ERA5-land data.

Chapter 4 addresses lake ice as a potential driver of trends and variability in algal biomass within study lakes over a 19-year period. This research explores the preprocessing of the ESA CCI Lakes product and assesses the trends and variance of various annual and seasonal chl-*a* concentrations (proxy of algal biomass), LSWT, and lake ice cover (LIC) parameters. This chapter is planned for submission to one of the following journals: *Frontiers in Water*, *Water*, or the *Journal of Great Lakes Research*.

Chapter 5 explores whether LIC is a predictor of algal biomass, using empirical and machine learning techniques. In addition, classification techniques using a random forest (RF) model were implemented to determine whether LIC can determine whether a given pixel can return the lake an abnormal algal biomass concentration. This chapter is planned for submission to one of the following journals: *Frontiers in Water*, *Water*, or the *Journal of Great Lakes Research*.

Chapter 6 expands upon the previous studies by assessing the interaction effects of many physical and atmospheric forcings of lakes on the dynamics of algal biomass, at a higher temporal resolution (daily – monthly scale). This chapter utilizes a High-Order Dynamic Gaussian Bayesian Network (HO-DGBN) to provide a Directed Acyclic Graph (DAG), to

simplify the interactions, and provide a dynamic context, as there is often a lagged response between an environmental trigger and an algal biomass response. This chapter is planned for submission to one of the following journals: *Frontiers in Water*, *Water*, or the *Journal of Great Lakes Research*.

Chapter 7 summarizes the findings of the previous chapters, identifies key challenges, and directs the potential for future work. The manuscripts are included in their original format and therefore may include some repetition in dataset descriptions, study areas, and methodologies between chapters.

Chapter 2: Algal biomass response to environmental change and its detection via satellite remote sensors

2.1 Introduction

The earliest recorded descriptions of algal blooms (AB) were made by Giraldus Cambrensis in 1188 (Reynolds & Walsby, 1975). After the discovery of microorganisms in 1665 by Robert Hooke and the work of Dr. Arthur Hassall in 1850 (Hassall, 1850), phytoplankton were defined by Victor Hensen, where AB were termed a "flowering of the waters" (Hensen, 1887; Whipple, 1899; Reynolds & Walsby, 1975). Light and temperature were known as the main growth drivers of ABs by 1899, where the effects of phosphates (Atkins, 1924) and land use (Lackey & Sawyer, 1945) were assessed by the early to mid-20th century. ABs became a significant issue within North America in the 1950's (Scott, 1955; Carmichael; 2008), where Lake Erie became known as the "Dead Sea of North America" by the 1960-1970s due to dissolved oxygen (DO) depletion (Ludsin *et al.,* 2001; Manivanan, 2008). ABs were associated with economic loss and health risks (Gorham, 1964; Reynolds & Walsby, 1975), the consequences of which prompted research by David Schindler in the Experimental Lakes Area of Canada, which had shown the impact of phosphorus on the growth of algae in a natural system (Schindler, 1974). Studies found that a reduction in phosphate-based detergents could reduce phosphorus (P) loading (Gakstatter & Allum, 1978; Maki *et al.,* 1984), where voluntary reductions began in the United States and Canada in the late 1970s and were banned by the detergent industry in 1994 (Litke, 1999).

P limitation success in reducing eutrophication/AB occurrence (Schindler *et al.,* 2016; Smith *et al.,* 2015) was short-lived, as AB occurrences increased globally by the turn of the 21st century (Crossman *et al.,* 2019). A 20-year time series (1998-2017) study of the Western Basin of Lake Erie had found an increase in the duration (days) and severity of annual cyanobacterial ABs (Sayers *et al.,* 2019). ABs have been observed to increase in frequency in recent decades (Hou *et al.,* 2022) where increases in intensive agriculture, reduction of wetlands/forests, increase in nitrogen-rich waste waters, and dissemination of invasive species are prevalent theories for the increase in ABs (Mitsch, 2017; Waters *et al.,* 2021). Global

6

climate change is considered a significant driver of ABs and the shift to toxin producing cyanobacteria (Paerl & Huisman, 2008; Salmaso *et al.,* 2018), which can result in neurological, hepatological, and hematological damage (Hilborn & Beasley, 2015). While some consider climate change a major driver (Paerl & Otten, 2013; Edlund *et al.,* 2022), others contest that anthropogenic nutrient loading is a mandatory prerequisite (Wang *et al.,* 2021; Chen *et al.,* 2014). Climate change impacts various environmental variables, which ultimately impact the lifecycle and growth rate of algae.

2.2 Climate Change Impacts on Lake Phenology

There is significant evidence that climate conditions are changing at an accelerated rate within the last 100 years compared to the past, driven by anthropogenic activities (Hallegraeff, 2010; Waters *et al.,* 2016). Lakes are considered sentinels of change, where the impacts of changes in environmental conditions are reflected in the health and productivity of lakes (Adrian *et al.,* 2009). Conditions such as shifts in air temperatures, more frequent and intense heatwaves/droughts, desertification, and more frequent/intense precipitation events trigger change in lakes such as: increased water temperatures, reduction in water levels, aeolian dust deposition, higher external/internal loading, reduction of lake ice duration, changes in mixing classifications, longer thermal stratification, methanogenesis, and acidification of lakes. Ultimately, such significant change to lake physical and biogeochemical structures can significantly affect the biota within the lake, leading to increased algal bloom potential.

2.2.1 Air Temperature Increase

Environmental conditions are changing at a heterogeneous and unprecedented rate, impacting freshwater lake physical and biogeochemical systems (Woolway *et al.,* 2020). Air temperatures have increased by 1.09°C from 1850 – 2022, due in part to the anthropogenic emission of various greenhouse gases (*i.e*., Methane (CH_4), Nitrous Oxide (N_2O), and CO_2) (Pörtner *et al.,* 2022). The rise in air temperatures coincides with a rise in freshwater temperatures, with an average increase of ~0.30 – 0.34°C per decade (Komatsu *et al.,* 2009; O'Reilly *et al.,* 2015; Yang *et al.,* 2019; Woolway *et al.,* 2022). Regional studies coincide with

global results in various regions: 0.26°C per decade in China (Xie *et al.,* 2022), 0.40 – 0.46°C per decade in western Russia (Filatov *et al.,* 2019), 0.23°C per decade in Austria (Niedrist *et al.,* 2018), 0.25°C per decade in Colorado, US (Roberts *et al.,* 2017), 0.40°C per decade in Europe (Stefanidis *et al.,* 2022), amongst others (Woolway *et al.,* 2019).

2.2.2 Heatwaves and Droughts Increase

In addition to increases in average annual temperatures, global studies have reported increases in temperature anomalies (Varotos *et al.,* 2013; Liu & Duan, 2018; Perkins-Kirkpatrick & Lewis, 2020). Heatwave intensity has increased from $1 - 8$°C per decade (1950 – 2014), where Northern Canada, Europe and Asia are the most affected (Perkins-Kirkpatrick & Lewis, 2020). Climate models predict double the frequency of severe drought conditions for Europe at a $2 - 3$°C global temperature increase (Samaniego *et al.,* 2018). The frequency of 100-year droughts is also anticipated to increase, particularly in the mid-west and Northern latitude regions (Zhao *et al.,* 2020). Increased drought and heatwave conditions increase the duration and severity of stratification and lower lake water levels (Zhai et al. 2023). Additionally, drought conditions result in dried soil surface conditions, limiting percolation potential and increasing surface runoff during rainfall events (Rakkasagi *et al.,* 2023).

2.2.3 Desertification Extension

The extension of drought conditions, along with anthropogenic terrestrial modifications, is a primary driver of desertification, primarily for arid, semi-arid, and semi-humid regions (Emadodin *et al.,* 2019; Huang *et al.,* 2020). Desertification increases the amount of suspended aeolian dust (Ajaj *et al.,* 2017), resulting in higher iron (Fe) deposition (O'Neil *et al.,* 2012). Atmospheric dust is photo-reduced to ferrous ions (Fe^{2+}) (Shi *et al.,* 2012), which is reoxidized to ferric ions (Fe^{3+}) in waters of a pH > 6, while waters with a pH < 6 will not oxide and remain soluble (Molot *et al.,* 2014). However, drought conditions are regional and not a global change, as various regions show heterogeneous change (Dai *et al.,* 2018).

2.2.4 Precipitation Changes

Climate models suggest an increase in global average precipitation from 1979 – 2020, with mean trends of 0.006 – 0.0102 mm day^{-1} per decade (Gu & Adler, 2023). Despite the average increase in precipitation, the change is highly heterogeneous, where various tropic and subtropical regions have seen reduced precipitation, while high latitudes showed an increase (Dai *et al.,* 2018). Regions that experience increased precipitation are also expected to see extreme drought conditions, as shifts in precipitation are primarily a result of increased intensity (Zhao *et al.,* 2020). Global land regions are expected to observe an increase in precipitation extremes by 32±8% – 55±13% than present by 2100 under varying climate scenarios (Thackeray *et al.,* 2022). Others have found that, while the frequency of intense precipitation is increasing, the magnitude is only increasing significantly for the northwestern hemisphere (Papalexiou & Montanari, 2019). Regional studies indicate an observed increase in heavy storms by 7%, 14% of very heavy storms, and 20% of intense storms for the US from 1908 – 2000 (Groisman *et al.,* 2004). A UK study returned similar results, with an observed increase in intense precipitation events from 1996 – 2009, with more intense events projected (Kendon *et al.,* 2014). Intense precipitation events prevent percolation, increasing surface runoff/flood risks, particularly when followed by drought events (Smith *et al.,* 2013). Arctic precipitation is expected to increase at a greater rate than previously projected, where there is an expected transition from snow to rain in the summer/fall seasons, impacting lake ice formation timings (McCrystall *et al.,* 2021).

Despite the changes to summer/fall conditions, it has been observed that winter temperatures are increasing at the highest rate, particularly for high-latitude regions (Hansen *et al.,* 2006; Houze *et al.,* 2019; Zhang *et al.,* 2019a). As a result of changes in water temperatures and precipitation patterns, lake ice phenology has observed a significant reduction globally (Woolway *et al.,* 2020).

2.2.5 Reduction of Ice Cover

Ice-free duration for Northern hemisphere lakes has increased from 0.63 days per decade over the past 150 years to an increase of 1.83 days per decade from 1975 – 2005

(Benson *et al.,* 2012), where ice duration trends were found to be six times faster (1992 – 2016) compared to any other previous quarter centuries (Sharma *et al.,* 2021). Regional studies have returned similar findings: 2.2 days per decade in the Great Lakes Basin/ Northern Wisconsin (Hewitt *et al.,* 2018), 1.2 days per decade in Northern Europe (Lopez *et al.,* 2019), 0.31 – 0.34 days per year in Northeastern Canada, 0.10 – 0.14 days per year in Northern Europe, 0.34 – 0.60 days per year in Alaska, US, 0.34 – 1.05 days per year in Central Siberia (Šmejkalová *et al.,* 2016), and 1.91 – 2.21 days per decade in China (Yao *et al.,* 2016). Similar to air temperatures, Northern latitude lakes have experienced higher ice loss and earlier break up rates compared to lower latitude lakes (Du *et al.,* 2017), where rate of change is often due to air temperatures, lake morphometry, elevation, precipitation, and cloud cover (Brown & Duguay, 2010). Lakes have been observed to change in ice cover classification, where perennial ice-covered lakes have begun to show seasonal melt (Obryk *et al.,* 2019; Sharma *et al.,* 2013), seasonal ice-covered lakes have transitioned to intermittent coverage, and intermittently covered lakes now experienceno winter ice coverage (Sharma *et al.,* 2019). The number of intermittently ice-covered lakes is expected to increase significantly under varying climate scenarios, where by 2080, the current ~15,000 intermittent ice-covered lakes are expected to increase between 35,300 and 215,600 (scenarios: 2 – 8°C global air temperature increase). Ice cover plays an important role in facilitating the duration of light exposure, the timing of lake mixing, and the duration of stratification.

2.2.6 Shift in Lake Mixing Classes

Lake mixing indicates that the hypolimnion and epilimnion in a lake have equalized temperatures and densities, allowing for a cycling of waters within the water column (Jewson *et al.,* 2009). Mixing supports and enables the transfer of macro and micronutrients from the benthos throughout the water column, increasing their availability (Stockenreiter *et al.,* 2021). Lakes are classified by the patterns and seasonality of mixing a lake experiences (Table 2.1 and Figure 2.1), which is controlled by the morphometry of the lake, the sources of inflow/outflow, precipitation, wind speed, cloud cover, etc. (Rocha *et al.,* 2022).

Table 2.1. Lake mixing classes as defined by Lewis (1983).

Mixing Class	Mixing Sub-Class	Definition
Polymictic	Continuous cold polymictic	Ice forming lakes with continuous mixing throughout the year. Typically does not see seasonal stratification.
	Continuous warm polymictic	Non-ice forming lakes with continuous mixing throughout the year. Typically does not see seasonal stratification.
	Discontinuous cold polymicitic	Ice forming lakes with irregular mixing events throughout the year. Typically does not see seasonal stratification.
	Discontinuous warm polymicitic	Non-ice forming lakes with irregular mixing events throughout the year. Typically does not see seasonal stratification.
Monomictic	Cold monomictic	Ice forming lakes with a single summer mixing event (typical at high latitudes with extensive ice cover). Stratified under ice (fall-spring).
	Warm monomictic	Non-ice forming lakes with a single winter mixing event. Stratified spring to fall.
Dimictic	-	Ice forming lakes with two annual mixing events during spring and fall seasons. Stratified in the summer and winter months.
Meromictic	-	Non-ice forming lakes which never mix and are perennially stratified.
Amictic	-	Ice forming lakes which never mix and are perennially stratified (perennial ice cover).

It has been observed that as a result of earlier ice breakup, earlier mixing occurs, which allows for longer open water seasons, resulting in the occurrence of summer stratification in arctic lakes, transitioning from a cold monomictic to a dimictic class (Vincent *et al.,* 2012). Historically, dimictic lakes have also been observed to transition to warm monomictic lakes, and even polymictic lakes to dimictic (Adrian *et al.,* 2009; Ficker *et al.,* 2017). A study by Woolway & Merchant (2019) found that of 600 globally distributed lakes, 100 will change in mixing class, of which 17 will transition from dimictic to warm monomictic. The shift to dimictic mixing ensures summer stratification, which can significantly affect the rate of internal loading and dissolved oxygen (DO) within the lake, significantly impacting the biota.
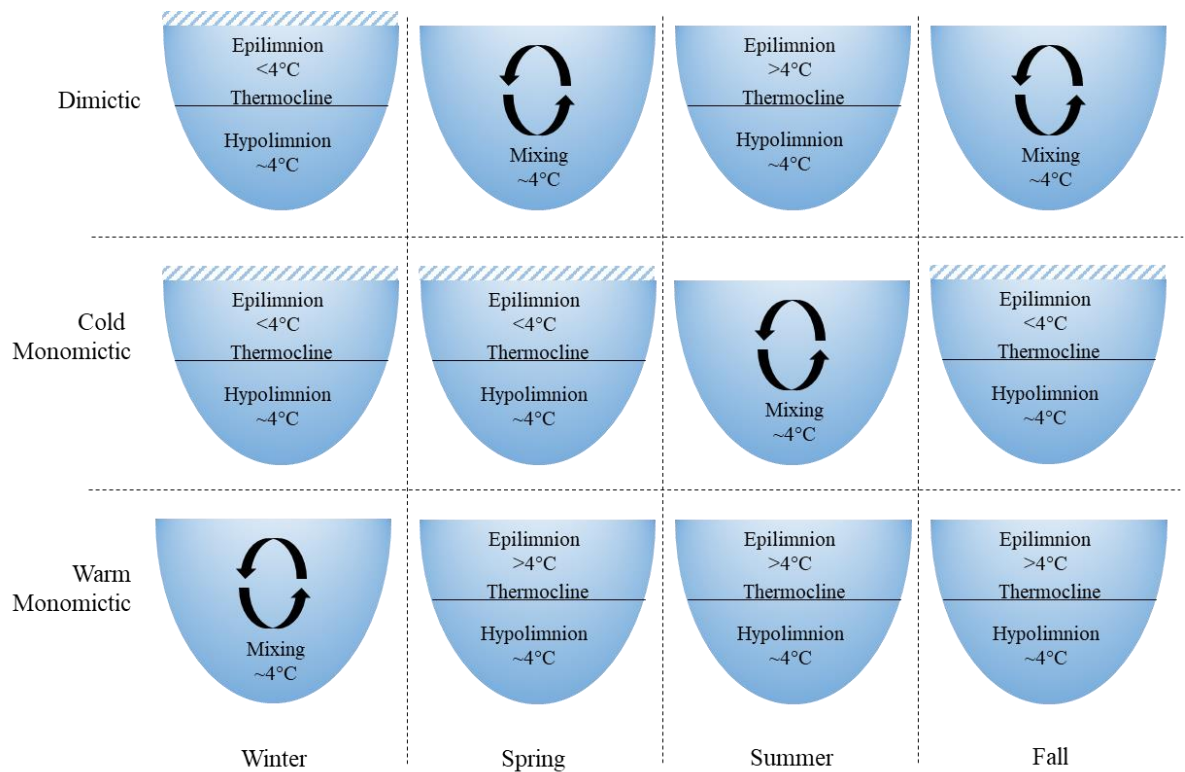
Figure 2.1. Examples of annual lake mixing classes (Dimictic, Cold Monomictic, and Warm Monomictic).

2.2.7 Increased Severity and Duration of Stratification

Due to the increases in global air temperatures, heatwaves, and droughts, the duration and severity of thermal stratification has increased in recent years for freshwater lakes (Kraemer *et al.,* 2015; Jane *et al.,* 2023; Wang *et al.,* 2023). Stratification occurs when prolonged solar radiative flux and air temperatures create a difference in water density and temperatures between the epilimnion and hypolimnion, preventing nutrient exchange (Anderson *et al.,* 2017). Lake morphometry and wind speed also play a role in the timing of stratification (Tanentzap *et al.,* 2008). Lakes express a heterogeneous pattern of stratification, particularly in large and deep lakes, which take longer to stratify, have shorter stratification periods, and show greater persistence of temperature anomalies (Woolway and Merchant, 2018). Under climate scenario models, it is anticipated that lakes globally see stratification occur 22.0±7.0 days earlier and dissipate 11.3±4.7 days later than currently seen (Woolway *et*

12

al., 2021). Under various climate projections, it is anticipated that 12% – 66% of NLs will shift to the same thermal regimes of low-latitude lakes by 2080 – 2099 (Maberly *et al.,* 2020).

Precipitation and wind events can initiate mixing, increasing internal loading within the benthos and external loading due to runoff (Weinke & Biddanda, 2019; Zhang *et al.,* 2022). The increase in turbidity from external and internal loading in surface waters increases the rate of stratification onset due to higher irradiance scatter in surface waters and limited penetration throughout the water column (Coats *et al.,* 2006; Butcher *et al.,* 2015). The longer stratification occurs, the more intense the stratification is (thinner thermocline, higher temperature difference between the epilimnion and hypolimnion), thereby further increasing its duration, as surface waters take longer to cool (Mesman *et al.,* 2021). Stratification is also known to play a role in methanogenesis, where extended periods of anoxia increase benthic release of CH_4 in eutrophic lakes (Bartosiewicz *et al.,* 2019; D'Ambrosio & Harrison, 2021). The prospect of methanogenesis may exceed the offset of carbon sequestration performed by algae (Skwierawski, 2022).

2.2.8 CO2 Deposition

The emission of CO_2 is considered the most significant greenhouse gas contributing to anthropogenic warming, increasing the deposition of CO2 in lakes (Rao & Riahi, 2006; Mustafa *et al.,* 2021). The deposition of CO_2 in waters results in the formation of inorganic carbon compounds; carbonic acid (H_2CO_3) and bicarbonate (HCO_3-) (Talling, 1976; Cole & Prairie, 2013). For waterbodies with low primary productivity (oligotrophic), the deposition of H_2CO_3 decreases the pH, while high productivity (eutrophic waters) results in an increase in pH due to carbon sequestration (Talling, 1976; Verschoor *et al.,* 2013). Acidification may be more likely to occur during the fall as algal levels decline, and more alkaline in the summer (Raven *et al.,* 2020).

There is significant evidence of shifts in environmental conditions globally, where waters are becoming warmer, precipitation events (drought and rain) are becoming more extreme and frequent, lake ice is reducing, lake mixing regimes are changing, and stratification

rates are increasing. Water chemistry profiles are changing due to increases in external loading, higher rates of internal loading, CO2 deposition, and reductions of DO in benthic waters. NLs are experiencing the greatest rate of change and, therefore, are at the greatest risk. These impacts will have a significant effect on algal growth cycles, biomass accumulation, and species compositions.

2.3 Impact of Changes in Lake Phenology on Algal Biomass Dynamics

Lakes are changing in response to the change in climate, due to shifts in physical and biogeochemical parameters (Figure 2.2). The impacts of these changes are dynamic and complex and vary both spatially and temporally. The impact of changes in various physical or atmospheric parameters differs depending on lake morphometry, algal community composition, anthropogenic influences, watershed terrain, etc. Various studies provide insight into how shifts in environmental parameters impact the dynamics of algal biomass.
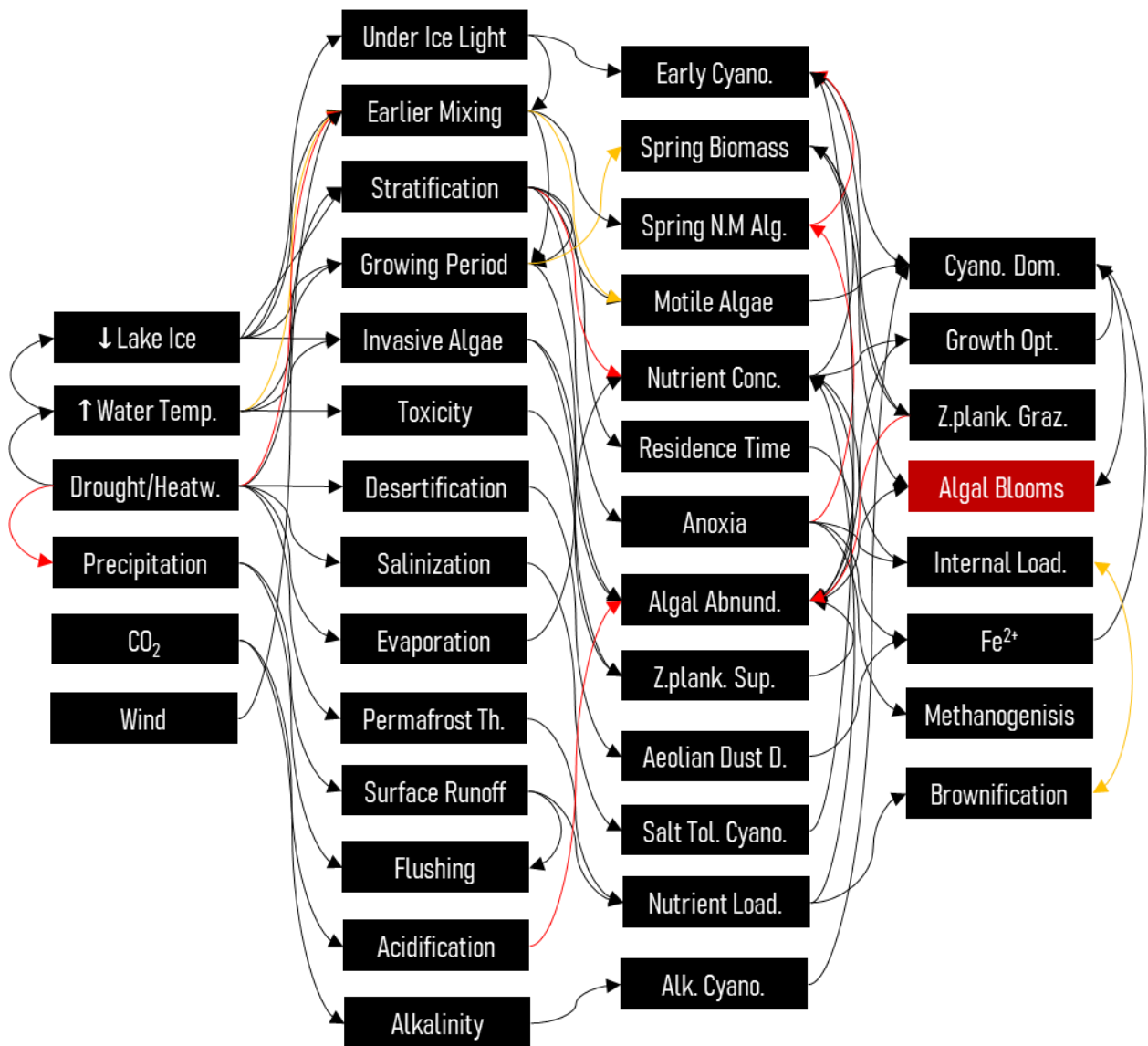
Figure 2.2. Pathways of biotic and abiotic drivers of algal blooms. The black arrows represent a positive interaction (an increase in the driver results in an increase in the target), red arrows a negative interaction (an increase in the driver results in a decrease in the target), and the orange arrows represent both a positive and negative interaction.

2.3.1 Impacts of Temperature on Algal Biomass

The increase in water temperature is frequently reported as the most significant climatic contributor to increased algal biomass accumulation, due to the impact warmer temperatures have on algal growth optimums (Paerl & Huisman, 2008; Paerl & Otten, 2013; Cross *et al.,* 2015; Borowitzka *et al.,* 2016; Thomas & Litchman, 2016; Gobler, 2020). Algal growth

optimums are genera specific, where various genera of cyanobacteria (*i.e*. *Anabaena* and *Microcystis*) proliferate at higher water temperatures (optima of ~30°C) while diatoms and dinoflagellates (optima of ~12 – 20°C) proliferate in cooler waters (Paerl & Otten, 2013). A study by Ho and Michalak (2019) showed that summer temperatures were a driving factor in algal biomass accumulation, and prolonged summer conditions drove cyanobacterial dominance in continental US lakes. Others had found that increased water temperatures are driving the dominance of toxin producers, where various cyanobacteria have been observed to express higher toxicity (Lamb *et al.,* 2019). NLs may also see an increase in cyanobacterial dominance due to the increase in lake thermal regimes supporting invasive species northward migration (Gobler, 2020). The increase in temperature for Northern regions results in a loss in permafrost, lowering albedo, and increasing regional radiance absorption (Chadburn *et al.,* 2017). Permafrost thaw is also linked to browning, increasing dissolved organic carbon (DOC), DOM and N, concentrations, impacting algal abundance and species composition (Kurek *et al.,* 2022; Myrstener *et al.,* 2022; Coleman *et al.,* 2023). The increase in nutrients drives primary productivity, which can increase autochtonous DOM, fueling primary productivity further (Miller *et al.,* 2009; Creed *et al.,* 2018). When primary productivity is high, however, surface waters become turbid and therefore waters become light limiting, reducing primary production (Gameiro *et al.,* 2011). The limitation of light penetration intensifies the rate of stratification in lakes.

2.3.2 Stratification Effects on Internal Loading

The increase in stratification may increase the growth rate of algae due to stable, non-turbulent water conditions, conditions that often favour cyanobacterial dominance (Brasil *et al.,* 2016; Lehman *et al.,* 2017). The accumulation of epilimnetic algae results in the eventual deposition of decaying algae in the benthos, where bacterial respiration increases, decreasing DO levels, where prolonged stratification prevents mixing, and further increases the severity of hypolimnetic hypoxia (DO $< 2mg\ L^{-1}$) (Nolan et al., 2019). The decomposition of algae in the benthos will also increase the deposition of nutrients, converting organic P to inorganic P, which can be made available during the following mixing cycle (Orihel *et al.,* 2017; Kim,

2018). Hypoxia may also increase the rate of hypolimnetic internal loading via lower redox potential (Woolway *et al.,* 2021). Sedimented Fe^{3+} is reduced under hypoxic conditions to Fe^{2+}, where low redox can break Fe-P complexes increasing P availability (Tammeorg *et al.,* 2020; Harrow-Lyle & Kirkwood, 2021). Stratified waters provide a competitive edge to cyanobacteria, which can migrate from the photic zone to the nutrient rich benthos, facilitating further growth (Huisman *et al.,* 2018). Cyanobacteria acquire Fe^{2+} for facilitating nitrogen fixation, where Fe^{2+} can limit cyanobacterial growth, however Fe^{3+} can be acquired via iron chelators (*i.e*., sideophores) (Molot *et al.,* 2014). Hypoxia may also indirectly support cyanobacterial dominance. Hypoxic conditions cause zooplankton to disperse into shallower waters, altering grazing patterns and reducing their efficiency, leading to suppression and potential mortality (Kimmel et al., 2009; Nolan *et al.,* 2019; Roman *et al.,* 2019). Poorer nutritional quality of cyanobacteria may also increase their dominance due to selective grazing by zooplankton (Caron & Hutchins, 2012; Senar *et al.,* 2019). Mixing can disrupt water column stability, allow aeration of the benthos and disperse benthic nutrients, providing advantages to non-motile algae such as diatoms (Sommer *et al.,* 2012; Jewson *et al.,* 2009).

2.3.3 Mixing Controls on Nutrient Availability

Mixing controls the timing of benthic nutrient suspension where macro- (P, N, C, and silicon (Si)), and micronutrients (Fe, copper (Cu), and zinc (Zn) facilitate algal growth and cellular reproduction, and are primarily found in their dissolved form (Raven, 1988; 1990; Smayda, 2008; Bonachela *et al.,* 2012; Song *et al.,* 2012; Lin *et al.,* 2016). P is considered the primary controlling nutrient of algal bloom formation (Schindler, 1974; Conley *et al.,* 2009). However, cyanobacteria have been hypothesized to benefit the most from N loading, as they can metabolize proteins from multiple N sources, both terrestrial and atmospheric (Zhan *et al.,* 2017; Erratt *et al.,* 2018). A lower N:P promotes cyanobacteria over other algae (Schindler *et al.,* 2008). Algae also require C for the metabolic process during photosynthesis, which can be sourced from DOM (comprised of ~50% DOC) or by atmospheric deposition (Talbot *et al.,* 1990; Zeebe & Wolf-Galdrow, 2001; Bade *et al.,* 2007; Moody & Worrall, 2017; Raven *et al.,* 2020). Micronutrients also play an important role in algal community compositions, where

higher Fe supports cyanobacteria due to its role in N_2 fixation (Hong *et al.,* 2017; Luo *et al.,* 2019; Freeman *et al.,* 2020). Higher Cu may also support cyanobacteria by increasing Fe bioavailability (Zhang *et al.,* 2019b), while Zn is also used for photosynthesis, it can become phytotoxic in large amounts (Ohki, 1976; Lahive *et al.,* 2011; Chen *et al.,* 2019a). Depending on the lake mixing structure, the timing and duration of mixing can impact the community composition and when algal biomass peaks occur, where cool turbulent waters will support non-motile algal growth (Salmaso, 2010). During winter, waters are typically stratified; however, turbulent waters provide earlier nutrient resuspension, therefore favouring several genera of dinoflagellates in spring phytoplankton assemblages over diatoms (Klais *et al.,* 2011; Hjrene *et al.,* 2019). The duration of mixing is important, as a short mixing cycle between the onset of mixing and stratification does not allow for full overturn, resulting in lower benthic DO, increasing internal loading, and has been known to lead to algal blooms, even for remote lakes (Favot *et al.,* 2019; Reinl *et al.,* 2021). Mixing is not only controlled by changes in air temperature, as strong wind and precipitation events can initiate mixing, primarily for polymictic lake systems (Woolway *et al.,* 2017; Holgerson *et al.,* 2022).

2.3.4 External Loading Driven by Precipitation

Increased wind speeds may induce mixing, stimulating algal growth for relatively shallow lakes, as wind-wave action can trigger mixing, suspending sedimented inorganic P (Huang *et al* 2016; Zhou *et al* 2021). Precipitation events allow for a mobilization of allochtonous sediments, often bound with nutrients, depending on the surface conditions of the catchment, where agricultural regions are often associated with high nutrient loading (Creed & Band 1998; Coffey *et al.,* 2019). Although increasing external loading, precipitation will induce flushing in lakes, disrupting stable growth conditions, depending on lake morphometry (Ho & Michalak, 2020; Bakker & Hilt, 2016). Lake waters become turbid following precipitation events, increasing the rate of stratification, where significant precipitation events followed by long periods of stratification and high temperatures are the predominant cause of cyanobacterial dominance (Paerl & Scott, 2010; Paerl, 2018; Coffey *et al.,* 2019). Rainfall events can replenish epilimnetic nutrients, prolonging/supporting cyanobacteria and higher

biomass (James *et al.,* 2008; Reichwaldt & Ghadouani, 2012); however, they may also disrupt/collapse high biomass and blooming events (Jacobsen & Simonsen, 1993; Jones & Poplawski, 1998; Ahn *et al.,* 2002; James *et al.,* 2008). Others have observed no significant correlation between precipitation and agal biomasses (Chinain *et al.,* 1999; Haakonsson *et al.,* 2017; Wiltsie *et al.,* 2018). Precipitation, along with temperature are significant contributors to lake ice formation timings, and ice type compositions (Ariano & Brown, 2019).

2.3.5 Lake Ice Controls on the Succession of Algae

Overwintering lakes were historically assumed ecologically dormant; however, under ice algal biomasses and even blooms have been observed (Hampton *et al.,* 2015; Block *et al.,* 2019). Waters under ice are typically stratified; however, snow depth and ice composition are the primary controllers of light penetration, which can induce mixing under ice (Palshin *et al.,* 2019). Highest algal biomass has been observed in ice-covered lakes without snow (Kalinowska *et al.,* 2019). Mixotrophs and diatoms favour the cold and dark water conditions present under ice and during spring conditions, where the typical succession of algae begins with spring biomasses dominated by diatoms and cryptophyceae, leading to green algae in early summer and cyanobacteria in late summer/fall (De Senerpont Domis *et al.,* 2007; Fang & Sommer, 2017). The reduction in ice cover has been observed to favour various dinoflagellates (*e.g. Ceratium hirundinella*) in spring assemblages (Salmaso *et al.,* 2018). In some cases, cyanobacteria have been observed to dominate spring assemblages when winter temperatures increase and light is not limited (Blenckner *et al.,* 2007; Cameron *et al.,* 2022; Rohwer *et al.,* 2023). The presence of lake ice controls the timing of mixing and the duration of stratification, indicating its importance in algal biomass assemblages throughout the year.

Climates are changing at a rapid rate, where the effects on algal biomass, bloom trends and community composition are complex. Lakes differ in their response due to their morphometry, legacy nutrients, climate conditions, etc. Higher temperatures increase algal growth rates to an optimum, the exceeding of which exhibits a negative effect. The duration of stratification is controlled by the duration of lake ice, wind, and precipitation events, which can affect the amount of internal and external loadings significantly impacting algal biomasses

depending on the community compositions. While there is an understanding of the important growth factors, forward models for bloom formation prediction have not been developed due to the inherent complexity of these interactions and lack of data (Janssen *et al.,* 2019). The application of remote sensing systems has allowed the use of historic image archives to provide comprehensive spatial and temporal coverage of algal biomass trends, potentially improving our understanding of these complex interactions.

2.4 Application of Remote Sensing for Algal Biomass Retrieval

To improve our understanding of algal biomass dynamics, remote sensing can be applied to extend the in-situ data record by providing approximations from satellite or aerial observations. Remote sensing provides instant observations at a wider spatial and temporal range, at the expense of limited accuracy due to sensor issues, turbid water conditions, over/underfitting of retrieval algorithms, and atmospheric correction issues (Mohebzadeh *et al.,* 2021; Mathews *et al.,* 2023). Freshwater lake monitoring is a considerable challenge, with an estimated ~117 million lakes globally at a size of > 0.002 km^2, *in situ* monitoring is often limited to non-remote, anthropogenically significant lakes (Verpoorter *et al.,* 2014). Due to the spatial extent of lakes, *ex situ* observations have become a popular method, as *in situ* sampling is lengthy, expensive and not always possible (Bosse *et al.,* 2019). Remote sensors detect differences in incoming/outgoing electromagnetic radiation (EMR), where incoming spectral irradiance (Esun) interacts with various surface and atmospheric molecular compounds with varying levels of absorption, reflection and transmission differing by wavelength, impacting the outgoing radiance (L$_\lambda$) (Jensen, 2005). Surface features can be detected by the difference in E$_{sun}$ and L$_\lambda$ at given wavelengths (from gamma (10^{-16} m) to radio waves (10^8 m)) measured with a spectroradiometer (Jensen, 2005). Studies as early as Boland (1975) have utilized remote sensors to detect algal biomass in waters. Since then, a wide variety of satellites have been applied to detect algal biomass.

2.4.1 Overview of Satellite Systems

There is a wide variety of remote sensors available, the applicability of which varies depending on the field of study. Remote sensors may be active or passive, each of which may be multispectral or hyperspectral. Active vs passive sensors differ in their method of data collection, where active sensors emit a signal for illumination, while passive sensors detect L_λ, primarily via solar illumination (Zhu *et al.,* 2018). Hyperspectral sensors measure L_λ at continuous intervals as small as 1nm, while multispectral sensors measure L_λ at irregular intervals that cover EMR ranges from 10nm to >100nm (Stuart *et al.,* 2019). Continuous measurements of hyperspectral sensors can improve surface feature identification; however, they are much more data intensive than multispectral sensors and have less applicability for historic data analysis and image archiving (Vincent *et al.,* 2004; Stuart *et al.,* 2019). Therefore, multispectral satellite sensors are better suited to long-term observations of algal biomass. However, satellites are limited by cloud cover and often have poorer spatial/temporal resolution due to their orbital path (Schott *et al.,* 2016; Rasti *et al.,* 2018). Satellite sensors vary in their spatial/temporal resolution, in which higher temporal resolution means larger swath width and lower spatial resolution (Schott *et al.,* 2016; Rasti *et al.,* 2018).

Satellite image quality and file size also depend on the bit-depth, which determines the number of unique observations the sensor records (e.g. a 6 bit system detects 64 (2^6) unique observations, while 12-bit can detect 4096 (2^{12}) observations) (Mika, 1997; Gerace *et al.,* 2013). The number of along-track detectors the sensor has may also improve the signal-to-noise ratio (SNR) by increasing their surface scanning duration (Vermote *et al.,* 1997; Schott *et al.,* 2016; Rasti *et al.,* 2018). Sensors with low bit-depth and SNR (*e.g*., Landsat 5 and 7) may be unable to discern between dark surface features and have lower applicability in oligotrophic waters (Schott *et al.,* 2016). A wide range of satellites have been used to retrieve surface water algal biomass concentrations, where optical satellites such as the Landsat satellite series (TM, ETM+ and OLI), MODerate resolution Imaging Spectroradiometer (MODIS), MEdium Resolution Imaging Spectrometer (MERIS), and the Sentinel satellite series (2A and

3A) are the most commonly used. The retrieval of surface water algal biomass must first minimize the impacts of atmospheric and optically active constituents (OACs).

2.4.2 Interactions of Light and the Atmosphere: Atmospheric Correction

Atmospheric aerosols (water vapour, aeolian dust, atmospheric gases, etc.) are highly heterogenous, impacting the rate and direction of irradiance scatter, and atmospheric pressure affecting the optical depth observed by a sensor (Vermote *et al.,* 1997). As atmospheric conditions vary, so too will the L_λ observed by the satellite (Song *et al.,* 2021). Freshwater systems are also highly heterogenous, where OACs such as algal and non-algal particles (NAPs) vary in concentrations, thereby impacting the inherent optical properties (IOPs) of the water (Werther *et al.,* 2022). The apparent optical properties (AOPs) observed by the sensor as surface remote sensing reflectance (R_{rs}), are dependent on the IOPs, and therefore vary significantly, where it is important to first address what the contribution of the atmosphere is before making assumptions of the surface water conditions (Pan *et al.,* 2022). Atmospheric correction over freshwater is, however, complicated and often a source of error in algal biomass retrieval algorithms.

The application of atmospheric correction can be either image or radiative transfer-based. A commonly used image-based correction is that of the Dark Object Subtraction (DOS) method (Vincent, 1972; Chavez, 1988; 1996), which assumes that all L_λ of shaded relief areas are due to atmospheric haze, representing a black body (black pixel assumption). Conversely, radiative transfer assumes a non-Lambertian surface and uses various atmospheric forward models to calculate two scatter models: Rayleigh (scatter by atmospheric gases), and Mie (scatter by atmospheric aerosols with a diameter $\geq$ intercepting wavelength of light) (Figure 2.3) (Vermote *et al*., 1997). Radiative transfer models such as the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) (Tanré *et al.,* 1990; Vermote *et al.,* 1997) and LibRadtran (Mayer & Kylling, 2005), however, include the black pixel assumption (Dark Dense Vegetation (DDV) model) for estimating aerosol optical thickness (Liu *et al.,* 2016; Müller-Wilm, 2017). Many early algal biomass retrieval algorithms had used the DOS method (Worth *et al.,* 1989; Ritchie *et al.,* 1990; Lathrop *et al.,* 1991; Carrick *et al.,* 1994), while

22

studies by the late 2000s commonly used radiative transfer correction methods (De Moraes Novo *et al.,* 2006; Li *et al.,* 2007; Wu *et al.,* 2009; Han *et al.,*2010). Both DOS and radiative transfer models violate the black pixel assumption, as freshwater lakes are often the darkest features for many satellite images, where L_λ signals are due to the IOPs and AOPs of water, and not just contributed by the atmosphere (Siegel *et al.*, 2000; Wang *et al.,* 2019). The violation of the black pixel assumption may introduce erroneous errors (*e.g*., anomalous or non-physical), particularly over turbid waters (Wang *et al.,* 2019). To avoid these errors, researchers have used simplified radiative transfer equations, correcting for only Rayleigh scattering, and retaining atmospheric aerosol contribution (Matthews *et al.,* 2012; Matthews & Odermatt, 2015; Zhang *et al.,* 2016; Tao *et al.,* 2017; Pahlevan *et al.,* 2020). Improved aerosol contribution models over lakes are a current and significant field of research for accurate retrieval of freshwater quality parameters (Molkov *et al.,* 2022; Pan *et al.,* 2022; Wang *et al.,* 2023; Zhao *et al.,* 2023;). Although atmospheric correction may introduce some errors, it is the differentiation of algal biomass signals from other OACs that is difficult.
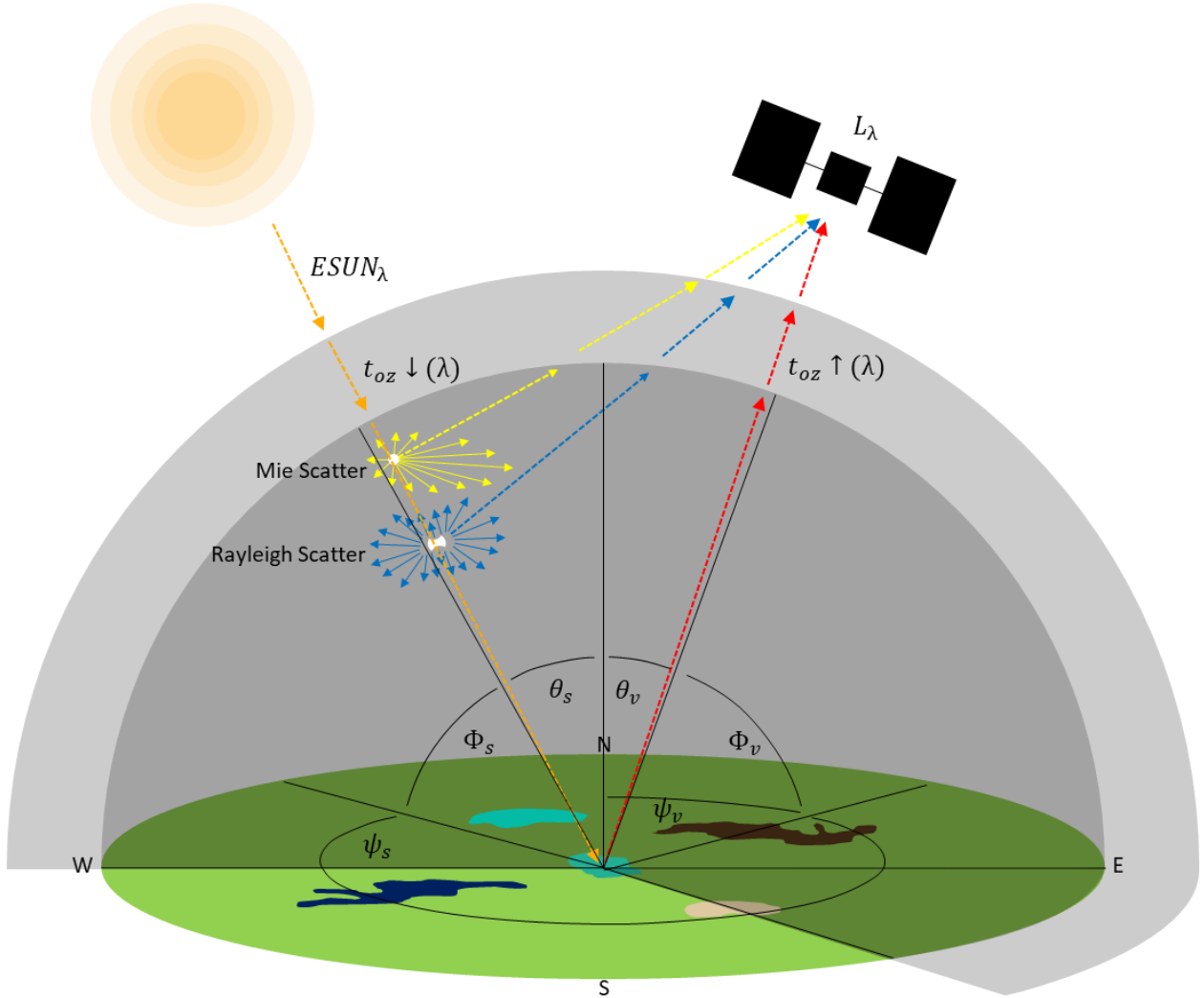
Figure 2.3. Geometry and pathways of visible-NIR atmospheric scatter observed by a satellite sensor where: $ESUN_\lambda$ is the exoatmospheric irradiance, $t_{oz}\downarrow(\lambda)$ is the downwelling ozone transmittance, $t_{oz}\uparrow(\lambda)$ is the upwelling ozone transmittance, L_λ is the observed at satellite radiance, θ_s is the solar zenith angle, θ_v is the satellite zenith angle, ϕ_s is the solar elevation angle, ϕ_v is the satellite elevation angle, ψ_s is the solar azimuth angle and ψ_v is the satellite azimuth angle.

2.4.3 Interaction of Light and Water: Complications by OACs

The diffusion of light at the air-water interface can vary by the geometry of light, surface water conditions (*i.e.*, waves), and the presence of OACs. Factors such as solar glint occur when the Esun is at a low solar elevation angle (Φ_s), increasing the observed L_λ (Dinnat & Le Vine, 2008). The surface water conditions impact the diffusion of light, as still waters cause specular reflection, while rough surfaces can result in Lambertian or non-Lambertian diffusion (De Keukelaere *et al.,* 2023; Yang *et al.,* 2023). Specular reflection is defined by

Snell's law, in which the solar zenith angle (θ_s) is equal to the sensor viewing angle (θ_v), while a Lambertian surface reflects the incoming E_{sun} in all directions (Reifsnyder, 1967; Bukata *et al.,* 1995; Yang *et al.,* 2023). Only a portion of incoming E_{sun} will reflect off the air-water interface, and the remaining light will be either absorbed or transmitted through the medium, resulting in subsurface scatter and reflection (Pinnel *et al.,* 2004). E_{sun} penetration in water is dependent on the concentration/composition of OACs and the wavelength (Figure 2.4) (Shanmuga Priyaa *et al.,* 2022). Marine waters have observed light penetration of ~1000m at 450 – 490nm (Reynolds & Lutz, 2001), however only up to ~55m at 475nm is observable by satellite sensors (Gordon & McCluney, 1975). In shallow, clear waters, light may reflect from the bottom of the lake increasing the observed L_λ, introducing error in retrieval algorithms (Bielski & Toś, 2022; Deutsch *et al.,* 2022). However, light penetration in freshwater lakes is highly variable, due to the significant variance of OACs.
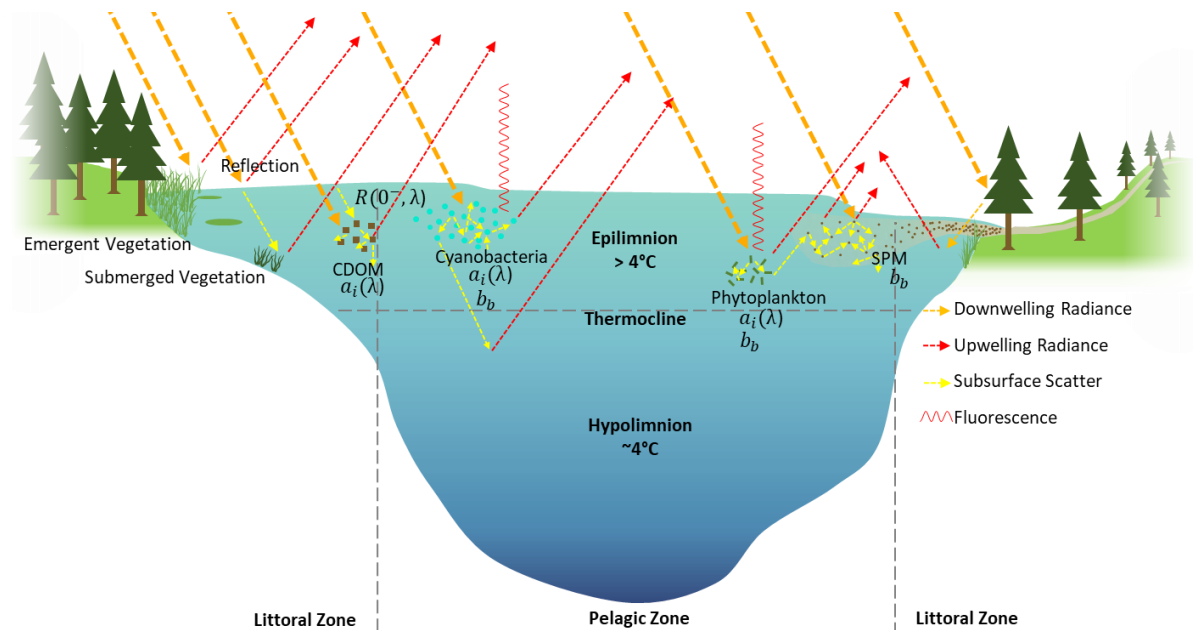
Figure 2.4. Freshwater optics for the remote sensing of phytoplankton where: CDOM is Colour Dissolved Organic Matter, SPM is Suspended Particulate Matter, $R(0^-, \lambda)$ is the subsurface spectral irradiance, $a_i(\lambda)$ is the absorption coefficient at a given wavelength, and b_b is the backscatter coefficient.

OACs refer to constituents present in waters that contribute to the L_λ observed by a sensor (the IOPs of the water), which include coloured dissolved organic matter (CDOM), non-algal particles (NAPs) and the photosynthetic pigments of algae such as: chlorophyll-*a* (chl-

a), chlorophyll-*b*, chlorophyll-*c*, phycocyanin, etc. (Xue *et al.,* 2019; Sagan *et al.,* 2020). The concentration and composition of OACs affect the rate of transmittance, absorption, scatter, and reflection of water; and are the primary source of variance for freshwaters (Devlin *et al.,* 2008; Soja-Woźniak *et al.,* 2020). The contribution of IOPs is determined by the absorbance of the matter and its backscatter, which can be modelled via Beer's Law, depending on wavelength, and the assumption that the attenuation length of light is proportional to the concentration of the medium (Smith & Baker, 1978; Bukata *et al.,* 1995; Carder *et al.,* 1999). These assumptions may be valid for pure water; however, in freshwater lakes, IOPs are in constant flux, making the establishment of forward models difficult (Janssen *et al.,* 2019; Moore *et al.,* 2017). For example, the absorption rate of chl-*a* observed by a sensor varies depending on the concentration and presence/concentration/composition of NAPs (Schalles, 2006). Chl-*a* exhibits an absorption peak at 440nm, coinciding with CDOM and NAPs; therefore, the presence of CDOM or NAPs may be confused with chl-*a* at these wavelengths (figure 2.5) (Giardino *et al.,* 2018). Recent research has implemented prior classification of optical water types (OWTs), to separate pixels by L_λ using supervised/unsupervised classifiers for the development of different algal biomass retrieval algorithms (Moore *et al.,* 2014; Spyrakos *et al.,* 2018; Neil *et al.,* 2019; Rodrigues *et al.,* 2022). Algorithms developed for differing OWTs havebeen found to improve algal biomass retrieval as the variance in IOPs is minimized (Dallosch & Creed, 2021; Dang *et al.,* 2023). Algal biomass retrieval algorithms use *in situ* observations in combination with L_λ to derive an inverse model, using empirical, bio-optical and machine learning techniques.
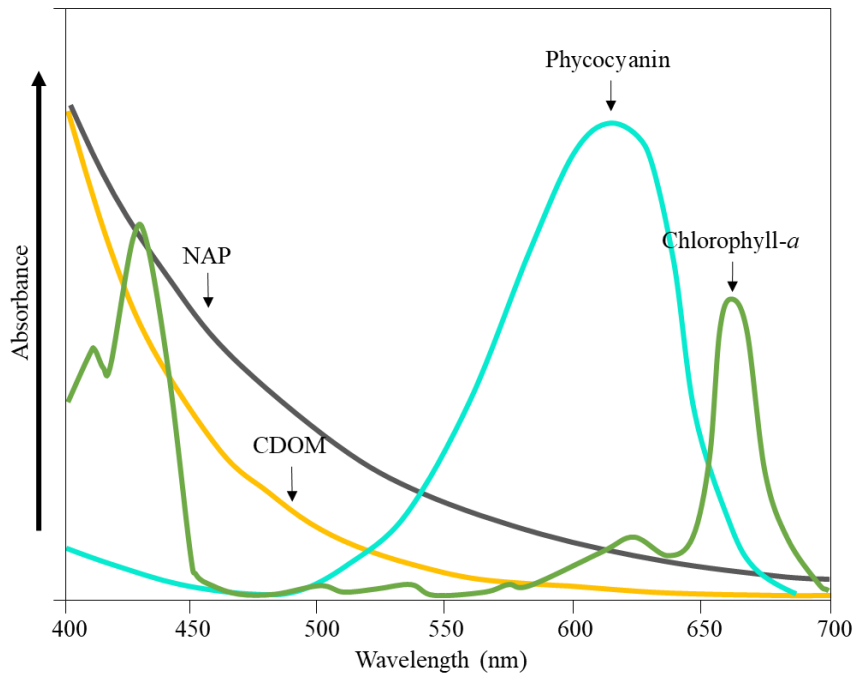
Figure 2.5. Absorption spectral curve of Chlorophyll-*a*, Phycocyanin, Coloured Dissolved Organic Matter (CDOM) and Non-*a*lgal Particles (NAPs). Adapted from: Roesler (2014) and Poniedziaek et al. (2017).

2.4.4 Empirical Regression Methods for Algal Biomass Retrieval Algorithms

Empirical regression techniques for algal biomass retrieval are one of the most frequently implemented due to their simplicity and low computation time (Odermatt *et al.,* 2012; Yang *et al.,* 2022). The empirical approach formulates a regression/multiple regression between L_λ and *in situ* observations of chl-*a* as a proxy of algal biomass (Ruiz-Verdú *et al.,* 2008). Regression methods utilize, *in situ* observations of surface water chl-*a* paired within ±1-7 days of L_λ observations at wavelengths that strongly denote algal presence, typically at the absorbance (440 and 670nm) and reflectance peaks (550nm and >700nm) (Schalles *et al.,* 2006). Hyperspectral sensors can specifically target these peaks, however, multispectral sensors measure L_λ at a wider spectral range, within fewer radiometric bands (Johansen *et al.,* 2023). Depending on the multispectral satellite, not all absorbance or reflectance peaks are covered, where wider bandwidths may also include other OAC peaks, significantly impacting the sensors' capacity to resolve chl-*a* (Li *et al.,* 2019). Multispectral sensors typically employ band mathematics, where various bands are used to minimize the impact of signal mixing, and

maximize the signal expected from chl-*a*, often targeting bands centered on the maximal absorption and reflectance of each parameter (Yang *et al.,* 2022).

Common empirical retrieval algorithms use a ratio of blue (0.44-0.5µm) or red (0.6-0.7µm) to green (0.5-0.6µm) or NIR centered bands (>0.7µm) (O'Reilly *et al.,* 1998; Sass *et al.,* 2007; Ha *et al.,* 2017; Salem *et al.,* 2017). Band ratios such as blue:green have shown high applicability in oligotrophic waters (Carder *et al.,* 2004; Sudheer *et al.,* 2006; Dang *et al.,* 2023), where the majority of the optical signal is that of water and chl-*a*, bands used in algorithms such as the ocean colour chlorophyll (OC) models developed by NASA (O'Reilly & Werdell, 2019). For turbid waters, a band ratio of the red-NIR edge (620nm; maximal refectance) to NIR bands (670nm; maximal absorption) has shown improved retrieval (Han *et al.,* 1994; Singh *et al.,* 2014), as longer wavelengths are less impacted by tripton and sediments due to the high absorption of water minimizing backscatter and increasing the chl-*a* fluorescence signal (Schalles, 2006; Matthews, 2011). The success of band ratios varies, as studies have found success with other ratios such as blue:red (Han & Jordan, 2005) and green:red (Brezonik *et al.,* 2005). More advanced band mathematics have also been employed, such as the three-band algorithms (Gitelson et al. 2003; 2011), normalized difference algal index (Ghunowa *et al.,* 2019), floating algal index (Page *et al.,* 2018; Rodríguez-López *et al.,* 2023), empirical orthogonal functions (Xi *et al.,* 2020), and line-height algorithms (Zeng & Binding, 2019; Tong *et al.,* 2022). The relationship between chl-*a* and L_λ may not be linear, where polynomial and quadratic equations have been applied, however, are prone to overfitting (Tiwari *et al.,* 2022). Machine learning methods have been used to address issues of non-linearity.

2.4.5 Machine Learning Methods for the Retrieval of Algal Biomass

Various machine learning methods have been used for the retrieval of algal biomass parameters such as support vector machines (SVMs), random forests (RF), and artificial neural networks (ANN). SVM methods include the use of regression, or SVR, which can classify high-dimensional data by defining a hyperplane, which is determined by non-linear mapping function via a kernel, which can be linear, polynomial, radial, etc. (Awad & Khanna, 2015).

28

The hyperplane defines the weights and bias to derive a relationship between L_λ of various radiometric bands and *in situ* chl-*a*, which has been used for algal biomass retrieval problems in previous studies (Kong *et al.,* 2017b; Kim *et al.,* 2014).

The RF method utilizes a series of regression trees that separate the inputs into clusters based on the total error observed within each cluster. For application in phytoplankton biomass retrieval by remote sensors, multivariate band inputs can be used as predictor variables $(x_1, x_2 \ldots x_n)$ to the response variable of chl-*a* or PC (y). Regression trees as described by Breimen *et al.*, (1984) use a series of linear regressions to determine predictor thresholds that produce the lowest possible error for the response variable. The predictor variable split points $(t_1, t_2 \ldots t_j)$ are binary and result in terminal nodes of the response variable ($Y_1, Y_2 \ldots Y_L$), where the binary split only stops due to a predefined criterion (*e.g.* minimum number of observations). Regression trees tend to overfit the training data, have low bias, and high variance, where other empirical measurement options are preferred. However, RF methods use many regression trees, where the input training observations are bagged, in which each tree consists of different training data. The process of bagging helps reduce variance in the testing data and show improved results compared to a single regression tree. Breiman (1996) describes the bagging process where the total data set ($\mathcal{L}$) consists of observation pairs $\{(x_n, y_n), n = 1, \ldots N\}$, where $y = \varphi(x, \mathcal{L})$ (in which $\varphi(x, \mathcal{L}_k)$ is a predictor) and is derived from multiple bags ($\{\mathcal{L}_k\}$) (in which k is the number of bags). When bagging, a randomized observation pair $\{(x_n, y_n), n = 1, \ldots N\}$ may either not be used at all in the model, or used more than once across all bags and even within the same bag (Breiman, 1996). RF methods have been used to derive chl-*a* concentrations from satellite data (Shi *et al.,* 2022; Xu *et al.,* 2022; Leggesse *et al.,* 2023); however, other machine learning approaches such as an ANN are also commonly implemented. For a more in depth look at bio-optical models, please refer to supplementary section A2.

The structure of an ANN is composed of 3 primary layers (Figure 2.6): input, hidden, and output, in which the hidden layer is represented by a complex set of weights and biases to determine node (or neuron) activation, mimicking the human brain process (Mani &

Srinivasan, 1997; Chang & Bai, 2018). ANN excel at identifying complex, nonlinear patterns, and typically require a high number of training observation pairs compared to other empirical methods, where the higher the number of inputs, nodes, and layers, the more complex the network and the longer the training time (Chang & Bai, 2018). Many ANN structures have been used for algal biomass retrieval, such as convolution (Nofrizal *et al.,* 2022; Xu *et al.,* 2023), or back propagation (Xu *et al.,* 2022; Zhu *et al.,* 2022;), however a common structure for regression purposes is the single layer feedforward neural network (SLFN), the structure of which use x,y pairs ($\{x_i, y_i\}$) as defined by (Chang & Bai, 2018):

$$\sum_{i=1}^{L} v_i g_i(x_j) = \sum_{i=1}^{L} v_i g(a_i \cdot x_j + b_i) \ (j = 1,2,3,\dots,N) \tag{2.1}$$

in which L are the total hidden nodes, j the total input samples, $a_i = [a_{i1}, a_{i2}, \dots, a_{iN}]^T (i = 1,2,3,\dots L)$ which are the weights ith hidden node, $v_i = [v_{i1}, v_{i2}, \dots, v_{iN}]^T (i = 1,2,3,\dots L)$ are the weights of the ith hidden node to the output, $g_i(x_j)$ the activation function and b_i the bias of each hidden node. The nodes activation is calculated via an activation function (*i.e*.., Linear, cosine, ReLU, etc.), based on the given weights and biases, calculating the output layer (Apicella *et al.,* 2021). Weights and biases are determined by the optimization function (*e.g*., stochastic gradient descent), allowing users to define parameters such as learning rate and momentum, to find the coefficients that return to lowest error (Tieleman *et al.,* 2017). The number of epochs determines how many times the SLFN is trained with the same inputs, while the batch size indicates the number of training data is used, and thereby the number of training iterations within each epoch (Chang & Bai, 2018). The higher the number of epochs, the longer the training time, where too many iterations can cause overfitting (Perin *et al.,* 2021). Several studies have used SLFN for algal biomass retrieval, with improved performance over standard linear regressions (Baruah *et al.,* 2001; 2002; Wang *et al.,* 2008; Song, 2011; Song *et al.,* 2012; Huo *et al.,* 2014; Chen *et al.,* 2017; Batur & Maktav, 2019). The performance of inverse models such as an ANN depends on the timing between *in situ* observations and R_rs observations, atmospheric interference, and presence of OACs (Syariz *et al.,* 2019; Dong *et al.,* 2020).

Attempts have been made to construct forward models using semi-*a*nalytical methods such as bio-optical retrieval. More in-depth examples of RF and ANN methods can be found in supplementary section A1.
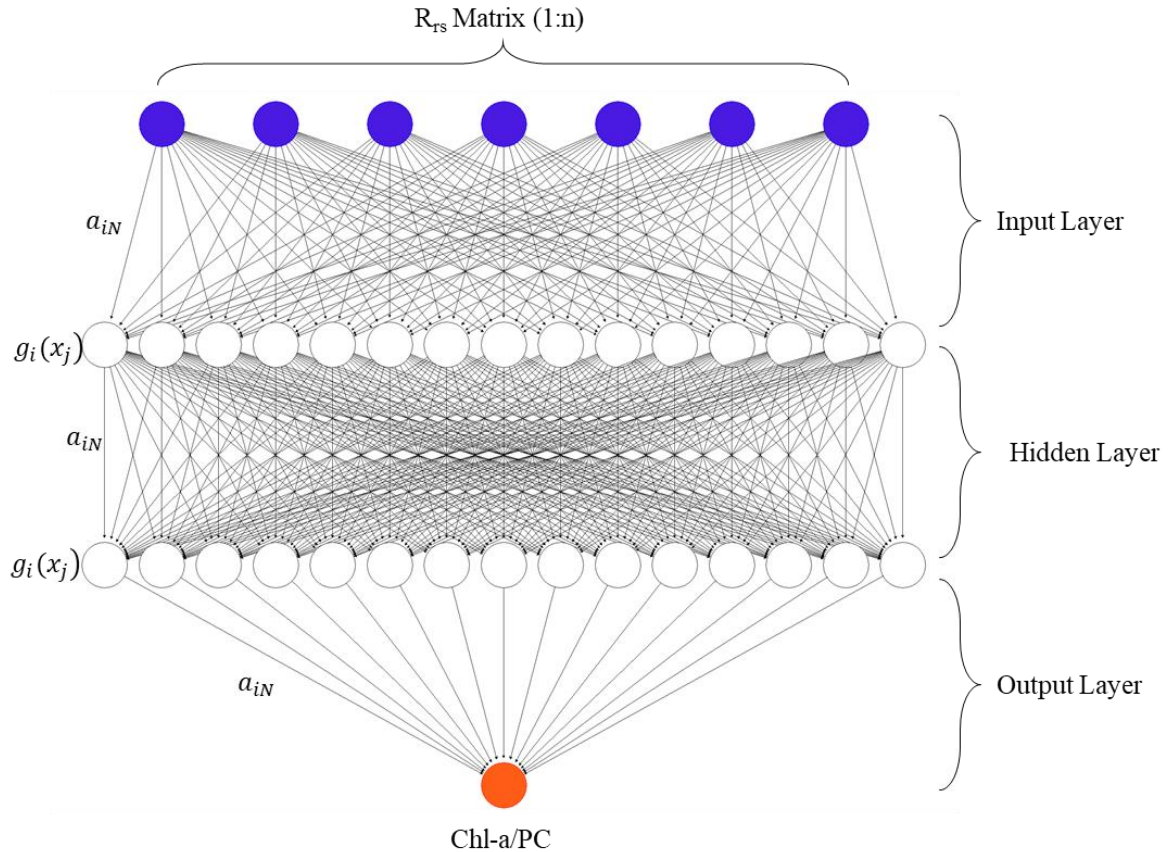
Figure 2.6. Artificial Neural Network Structure for chl-*a*/PC retrieval. $\boldsymbol{a_{iN}}$ represents the weights, $\boldsymbol{g_i(x_j)}$ represents the activation function.

2.4.6 Bio-Optical Methods for the Retrieval of Algal Biomass

The development of forward models for algal biomass retrieval in freshwater lakes is difficult due to the spatial-temporal change in IOPs, as an inversion model of AOPs and IOP to Rrs are required (Ogashawara *et al.,* 2017). Case-I waters (oligotrophic) have forward models developed (Morel & Prieur, 1977), however case-II requires an inversion. IOP inversion uses Rrs to derive an absorption coefficient at a given wavelength ($a_i(\lambda)$) of water ($a_w(\lambda)$), chl-*a* ($a_{Chl}(\lambda)$); along with the $b_b(\lambda)$ of water (Simis *et al.,* 2005). The change in the turbidity of water is proportional to the ratio of $a(\lambda)$ and $b_b(\lambda)$, as calculated by the Kubelka-

31

Munk remission function (Ogashawara *et al.,* 2017). The inversion of IOPs from Rrs assumes that the observed absorption is due to $a_{water}(\lambda)$, $a_{phyt}(\lambda)$, the absorption of tripton (or other NAPs; $a_{NAP}(\lambda)$) and the absorption of CDOM ($a_{CDOM}(\lambda)$), with $b_b(\lambda)$ a product of $b_{bwater}(\lambda)$ and $b_{bparticles}(\lambda)$ (Roesler & Perry, 1995). The spectral shape of each parameter is described by Beer's law; as the concentration increases, so does the IOP, which is predicted by the $a(\lambda)$ and *in situ* concentration (Smith & Baker, 1978; Roesler *et al.,*1989). Chl-*a* retrieval (proxy of algal biomass) can be calculated based on these inversions, as seen by Gons (1999), and later adapted for satellites such as MERIS and Landsat with high success (testing $r^2 = 0.68 – 0.96$) (Gons *et al.,* 2002; Simis *et al.,* 2005; Allan *et al.,* 2015). Studies have found higher inversion precision compared to empirical (Allan *et al.,* 2015; Kong *et al.,* 2015). The complexity and variability of IOPs in lakes can make the definition of the spectral shapes difficult, where models are often lake specific (Eleveld *et al.,* 2017). For a more in-depth look at bio-optical models, please refer to supplementary section A3.

2.4.7 Overview of Bayesian Networks Application

This research makes use of time-series analytics to infer the interactions within a multivariate system as they change over time and affect algal biomass dynamics. Any given time series may be composed of three given signals: the stationary, non-stationary and the residuals (Kulshreshtha, 2020). A stationary signal denotes the presence of seasonality within the data, whereby the mean and variance do not change over time (McElroy, 2012). The non-stationary signal denotes the presence of a trend, where the mean and variance change with time (Hammond & White, 1996). The residuals of a time series denote the noise and error, in which the change over time is random (Gujer, 2008). Statistical analysis often makes an assumption of homogeneity and normality, where the data are assumed stationary and normally distributed, however, this is rarely true with real-world data and are often left unchecked (Hoekstra, *et al.,* 2012).

To better understand the dynamics of algal biomass response in a system of varying atmospheric and lake physical variables, a Bayesian network (BN) analysis is preferred over

standard regression methods. Regression methods (*e.g*., Ordinary Least Squares) utilize a frequentist approach, in which model coefficients are fixed and constant across repeatable samples (Law *et al.,* 2014). BNs however, fit probabilities or distributions based on a fixed sample, where the parameters are random (Law *et al.,* 2014), and therefore are considered more flexible (Ickstadt *et al.,* 2011; Rhodes, 2017; StataCorp, 2015). Given that the parameterization of frequentist and BN methods vary, so do the approximations, both on a technical and theoretical basis. The frequentist approach constructs confidence intervals for each predicted value for given a hypothetical number of samples, where 95% of the intervals constructed would contain the true value (Morey *et al.,* 2016). Conversely, with a BN approach, credible intervals (posterior probability intervals) are constructed for each forecasted observation, in which there is a 95% probability that the value lies within (Flor *et al.,* 2020). Therefore, in regards to hypothesis testing, assuming that a given null hypothesis is true, a frequentist approach looks to understand how likely you are to observe your data. Instead, a BN approach asks how likely the null hypothesis is given your provided data (Rhodes, 2017; StataCorp, 2015). BNs offer some distinct advantages, such as the incorporation of priors (Samaniego, 2010), are considered more comprehensive due to the use of the entire posterior distribution returning direct uncertainty (van de Schoot *et al.,* 2013), can determine complex interdependencies between variables and can easily predict for multiple explanatory variables (Jackson-Blake *et al.,* 2022).
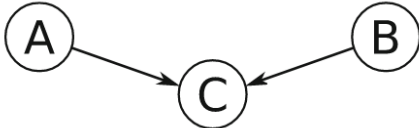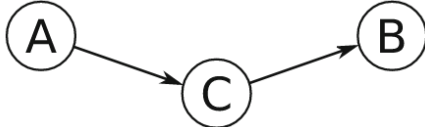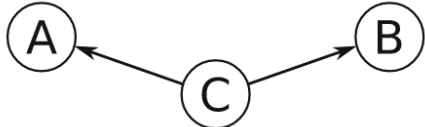
Other machine learning methods such as an Artificial Neural Network (ANN) and Random Forest (RF) are often ill-equipped for determining variable interdependencies, as they frequently take a "black box approach" (Koushik *et al.,* 2020). The increase in available data size and scope has led to a large-scale black-box optimization problem, where it becomes increasingly difficult to understand which decision variables returned the desired result (Chen *et al.,* 2022). For these reasons, the BN approach is most suitable for research goals in understanding the dynamics of algal biomass in freshwater lakes.

This research specifically implements a Dynamic Gaussian Bayesian Network (DGBN), however, to better understand its utility in achieving our research objectives, a typical discrete BN will be first described and built upon from there.

2.4.7.1 Discrete Bayesian Networks

A BN is trained using a set of discrete data amongst multiple variables and comprised of two components: a directed acyclic graph and a probability distribution (Sebastiani *et al.*, 2005). The goal of the BN is to identify the interdependencies of the input variables by calculating the conditional dependence of all variables. A variable with no dependency is considered a parent node in the network and is represented by a marginal probability table. Variables with conditional dependence are considered a child node and are represented as a joint probability table. Therefore, the state of the child node (x_i) is conditionally dependent on the state of the parent nodes (A_i), which is described as $P(x_i|A_i)$. A simplified network with three given nodes $P(A, B, C)$, may represent one of three fundamental connections (Table 2.2). A network may be comprised of one or all three connection types.

Table 2.2. Graphical separation indicating the conditional independence and probability decomposition of a three-node network. Modified from Nagarajan *et al.*, 2013.

Connection	Network	Equation	EQ (#)
Converging		$P(A, B, C)$ $= P(C|A, B)P(A)P(B)$	2.3
Serial		$P(A, B, C)$ $= P(B|C)P(C|A)P(A)$	2.4
Diverging		$P(A, B, C)$ $= P(A|C)P(B|C)P(C)$	2.5

A BN follows the Markov assumption, in which a node is conditionally independent of nodes that do not descend from it, while a child node is conditionally dependent on its parent nodes (Nagarajan *et al.,* 2013). The joint probability distribution of the variables within the network is enabled by the Markov assumption, in which the joint probabilities are a product of conditional probability distributions as described by the following equation by Nagarajan et *al.* (2013):

$$P_X(X) = \prod_{i=l}^{p} P_{X_i}(X_i | \Pi_{X_i})$$
(2.2)

Where X_i is a given variable and Π_{X_i} is the set of parents of X_i. A set of connected parent and child nodes that are separate from the rest of the network are considered a Markov blanket. As seen in Table 2.2, each ordered or unordered pair of nodes are connected via a directed arc (or vertex). Various methods can be implemented to determine the optimal structure of the network and the direction of the nodes. A constraint-based approach calculates all possible combinations of conditional independence tests and, although effective, is limited to smaller networks with a lower number of variables (Bretta *et al.,* 2018). Conversely, the network can be trained using a score-based approach, where the network structure is determined by maximizing the likelihood of the data estimation (Bretta *et al.,* 2018). More recent methods have utilized a heuristic search technique that seeks to minimize the BN search space. Many heuristic methods may also be considered score-based. Hybrid methods seek to optimize the network structure using both constraint and score-based methods. A detailed look at the different optimization structures can be seen in Table 2.3. The network now has a given structure, organized between parent and child nodes, where the state of the child node is a conditional probability of the state of their parent. Therefore, scenarios can be derived from the parent nodes to determine the state of the child node.

Table 2.3. Summary of fundamental Bayesian network optimization structures as described by Nagarajan et. al. (2013).

Type	Algorithm	Description	Source
Constraint-Based	PC	Uses a backwards selection procedure.	Sprites *et al.,* 2001

Type	Algorithm	Description	Source
	Grow-Shrink (GS)	Simple forward selection Markov blanket approach.	Margaritis, 2003
	Incremental Association (IAMB)	Forward selection followed by a backwards selection.	Tsamardinos *et al.*, 2003
	Fast IAMB	Reduces the number of conditional independence tests from IAMB using a speculative stepwise forward selection.	Yaramakala and Margaritis, 2005
	Interleaved IAMB	Avoids false positives of the IAMB through the use of a forward stepwise selection.	Tsamardinos *et al.*, 2003
Score-Based	Greedy Search (GS)	The search space starts with an empty graph, and iteratively changes (adds/removes) from the structure until the score no longer improves.	Bouckaert, 1995
	Genetic	Search space provides random alterations (mutations) to a combination of two network structures.	Larrañaga *et al.*, 1997
	Simulated Annealing	Stochastic local search which allows both an increase and decrease to the network score.	Bouckaert, 1995
Heuristic	Hill-Climbing (HC)	Local search space where incremental changes are made to decrease network error and find the local minimum.	Nagarajan *et al.*, 2013
	Tabu Search (TS)	Local search space which implements an adaptive memory to retract non-improving moves in finding the local minimum.	Glover, 1989
	Particle Swarm Optimization (PSO)	Local search space which implements many simultaneous 'particles' to search for the local minimum and updates the velocity after each iteration.	Sahin *et al.*, 2004
	K2	Local search to determine the local minimum given a user defined prior and implements a GS method.	Cooper & Herskovits, 1992

Type	Algorithm	Description	Source
Hybrid	Max-Min Hill-Climbing (MMHC)	Expansion of HS which minimizes the search space by using a smaller set of related nodes, which then seeks to maximize the network score.	Tsamardinos *et al.,* 2006
	Sparse Candidate	Similar to MMHC, however, the min and max steps are applied iteratively until the network score no longer improves.	Friedman *et al.,* 1999

There are several ways in which the BN can be used to answer questions (*e.g.* probabilistic reasoning or belief updating) a user may have regarding the system, known as inference (Pearl, 1988; Nagarajan *et al.,* 2013). Probabilistic reasoning utilizes either hard (instantiation of one or more variables) or soft (new distribution for one or more variables) evidence to determine the change in the network probabilities under a given hypothesis testing problem (Nagarajan *et al.,* 2013). Exact inferencing applies Bayes' theorem in conjunction with local computations to obtain exact probability values of a given target node, however, are often limited to small/simple networks (Nagarajan *et al.,* 2013). Approximate inference algorithms instead estimate the local probabilities using a Monte Carlo simulation and are therefore more applicable to high-dimensional networks where the multivariate distribution is complex (Nagarajan *et al.,* 2013; Kwisthout, 2018). Casual inference methods look to determine the change in the network based on the effects of interventions (separate dataset modified by the user). A summary of various inference methods can be found in Table 2.4.

Table 2.4. Summary of Bayesian network inferences as described by Nagarajan et al. (2013).

Type	Method	Description	Source
Probabilistic Reasoning	Conditional Probability	The change in the target node(s) is assessed given a change in the parent nodes, which is useful in experimental design factors or outcome analysis.	Nagarajan *et al.,* 2013
	Maximum a Posteriori	Looks to find the values of the target node(s) that have the highest posterior probability, which is useful for imputing	Nagarajan *et al.,* 2013

Type	Method	Description	Source
		missing data or comparing the true values with the new values of the target node(s)	
Exact Inference	Variable Elimination	Direct use of the BN structure, where the optimal operative sequence on the local distributions is used on the local distribution.	Nagarajan *et al.,* 2013
	Junction Trees	The network graph is converted into a tree, clustering the original nodes.	Nagarajan *et al.,* 2013
	Message-Passing	Inference on junction trees in which directed messages are passed from a node and its neighbors to the child node.	Korb & Nicholson, 2010
Approximate Inference	Rejection Sampling	Monte Carlo estimation of a complex distribution where the samples are derived from an approximate distribution, and only accepted given the approximation is ≤ true distribution; otherwise, the sample is rejected.	Robert & Casella, 2009
	Importance Sampling	Monte Carlo estimation of a complex distribution where the samples are derived from a modified distribution of the weighted average of random samples.	Robert & Casella, 2009
	Particle Filtering	Monte Carlo estimation computed by representing the posterior probability through individual sample generation and randomly sampling them based on their likelihood weighting. Useful for noisy data.	Moral, 1997; Orhan, 2012
Casual Inference	Ideal Interventions	Measures the effects of fixed interventions on the casual network structure as opposed to a belief change in the observed data.	Nagarajan *et al.,* 2013
	Stochastic Interventions	Similar to ideal, however, the interventions are not fixed and instead random.	Nagarajan *et al.,* 2013

Discrete BN have been applied in various fields such as healthcare (McLachlan *et al.,* 2020; Kyrimi *et al.,* 2021), economics (Gemela *et al.,* 2001; Masmoudi *et al.,* 2019), risk management (Khakzad *et al.,* 2013; Ruiz-Tagle, 2022), genetics (Vignes *et al.,* 2011; Su *et al.,* 2013), and environmental (Aguilera *et al.,* 2011; Barton *et al.,* 2012) modelling. Despite the wide applicability, such network structures are limited to discrete datasets, limiting their potential when only continuous measurements are available. If a network analysis is desired when only continuous data is available, as is often the case in economics and environmental modelling, a Gaussian Bayesian Network (GBN) can be implemented.

2.4.7.2 Gaussian Bayesian Networks

A GBN is structured the same as a BN, in which it is composed of a DAG, however, all random variables are continuous and assumed Gaussian distributions. Therefore, all parent nodes are an assumed Gaussian distribution, and all child nodes are a conditional multivariate Gaussian distribution, as opposed to the marginal and joint probability tables of the discrete BN (Grzegorczyk, 2010). The GBN therefore assumes a univariate normal distribution within the local network and a multivariate normal distribution in the global network (Neapolitan, 2004; Nagarajan *et al*., 2013). However, some studies have allowed local networks to violate this assumption utilizing non-parametric BN (Vitale *et al.,* 2020; Marella *et al.,* 2019). The parent-child node relationship is therefore represented as a linear Gaussian model as defined by Quesada *et al*. (2021a):

$$p(x_j|Pa_j) = \mathcal{N}(\beta_{0j} + \beta_{1j}x_{1(j)} + \cdots + \beta_{rj}x_{r(j)}; \sigma_j^2) \tag{2.6}$$

Where x_j represents the target node, $Pa_j = \{x_{1(j)}, \dots, x_{r(j)}\}$ is the set of parent nodes for x_j, $\mathcal{N}$ represents the density function of a normal distribution, β_{0j} is the independent coefficient (intercept), $\{\beta_{1j}, \dots, \beta_{rj}\}$ are the parent coefficients (slope), while σ_j^2 is the variance (error) of x_j. The parent-child node connections are represented by multivariate Gaussian linear regressions, however, this method differs from the frequentist approaches both as described in Section 2.4.4-6, and through the method of inference. Provided that all nodes are linear

Gaussian models, the joint Gaussian distribution of the network can be written as described by Quesada *et al.* (2021a):

$$p(x) = \prod_{i=1}^{n} p(x_j | Pa_j) = \prod_{i=1}^{n} \mathcal{N}(\beta_{0j} + \sum_{j=1}^{r(i)} \beta_{ji} \, x_{j(i)}; \sigma_j^2) \qquad (2.7)$$

Where $x = (x_1, \dots, x_n)$ represents the number of nodes within the network, and $r(i)$ represents the total amount of parent nodes for x_j. The model coefficients are estimated using a Gaussian multivariate linear regression distribution. Optimization and structure of the models is the same as a discrete BN, where any measure of metrics can be used as seen in Table 2.3. Once the network is constructed, the network can be used to inference new data based on the distributions of the parent nodes (Table 2.4).

Like discrete BNs, GBNs have also been applied to a variety of fields, including medical (Huang *et al.,* 2011; Wang *et al.,* 2020), economics (Charfi *et al.,* 2020), risk management (Arfin & Yodo, 2021; Liu *et al.,* 2023), genetics (Liu *et al.,* 2020; Graafland & Gutiérrez, 2022), environmental (Zhang *et al.,* 2012; Meineri *et al.,* 2015), and algal biomass (Jackson-Blake *et al.,* 2022; Deng *et al.,* 2023) modelling. There are instances where both discrete and continuous variables are needed, in which case a mixed discrete-continuous BN (MBN) is used (Hu *et al.,* 2022). Mixed data networks function the same as the discrete and continuous case. The MBN are inferenced as a conditional linear Gaussian (CLG) network, where discrete variables may only act as a parent node and not a child-node (Bøttcher & Dethlefsen, 2003; Hu *et al.,* 2022), however, CLG networks can be modified for inferencing discrete data (Lerner *et al.,* 2013). Discrete, continuous, and mixed BN are all considered static models, as they do not account for the temporal domain (Nagarajan *et al.,* 2013), where those that do are known as a Dynamic BN (DBN). This research makes use of only continuous data, and therefore only a Dynamic Gaussian Bayesian Network (DGBN) is described in section 2.4.7.3.

2.4.7.3 Dynamic Gaussian Bayesian Network

The DGBN is a desired network analysis method when concerned with the interaction of variables as they change over time (Quesada *et al.*, 2021). The structure of a DGBN is similar to that of the GBN; however, each variable is described in the temporal domain $X = \{X_i(t); i = 1, \dots, k; t = 1, \dots, t\}$. Variables are also conditionally dependent on the past observations of other variables by a given time-step (Nagarajan *et al.*, 2013). A time-step (lag interval) is measured at the temporal frequency of the input data (*i.e.*, daily observations where t_1 represents a 1-day lag, t_2 represents a 2-day lag, etc.). The time step indicates which ordered pairs of variables are represented in the multivariate linear Gaussian equation. For example, a network in which variable X_1 at t_1 is paired with the same variable X_1 at t_0, the ordered pair is of $t_1: t_1 - 1$. The joint probability of the DGBN now includes all time slices t to a given horizon T as described by Quesada *et al.* (2021a):

$$p(X_0, X_1, \dots, X_T) = P(X_{0:T}) = p(X_0) \prod_{t=0}^{T-1} p(X_{t+1} | X_{0:t}) \tag{2.8}$$

Where $X_t = (X_{1t}, X_{2t}, \dots, X_{nt})$ is the vector of all the nodes in a time slice t, where $t = 0, 1, \dots, T$ and T is the maximum number of time slices. There are four main assumptions of the DGBN as described by Nagarajan *et al.* (2013):

1. The stochastic process is first-order Markovian (*i.e.*, any variable at time t is dependent only on the past variables at time $t - 1$).
2. Variables at the same time t are conditionally independent given their immediate past.
3. The temporal profile of a variable cannot be written as a linear combination of the temporal profile of any other variable (*i.e.*, each variable's temporal profile is unique).
4. Often assumed the temporal process is homogeneous and invariant (*i.e.*, the observed phenomena are under the same rules for the entire study period).

Interaction effects can only be observed at the minimum temporal scale of the input data; therefore, any effects at a smaller scale cannot be observed in the network (*i.e.*, daily data inputs cannot account for hourly effects). Assumptions 1-3 are required for constructing an

accurate DGBN, however, assumption 4 is often not true with most real-world data and instead used to simplify the model (Nagarajan *et al.,* 2013). Flexible models that relax assumption 4 typically require a higher number of observations (Xuan Vinh *et al.,* 2012). Studies have successfully implemented DBGN with non-homogenous data (Grzegorczyk, 2016; Quesada *et al.,* 2021a). A first-order Markovian network as described by Koller and Friedman (2009) is one in which the state of a given time slice is dependent on only the preceding time slice (figure 2.7), which is important to understand given assumptions 1 and 2.
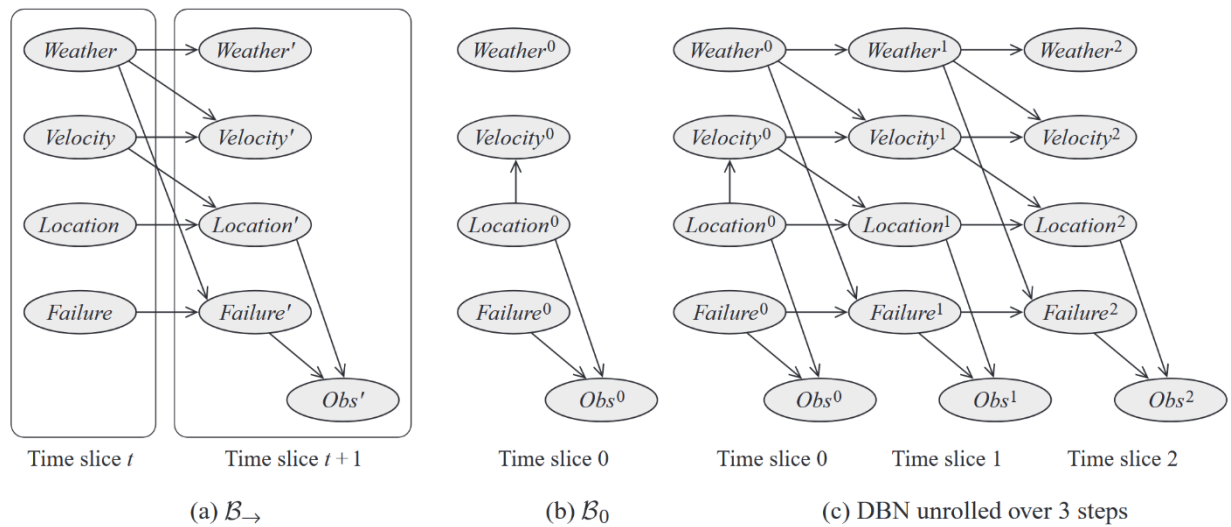
Figure 2.7. An example of a simplified DGBN where (a) represents a two-time slice, (b) represents a DGBN with 0 time slices, and (c) represents a DBN with 3 time slices. Taken from Koller & Friedman, 2009.

The joint distribution of a first-order Markovian model can then be summarized as the following equation as described by Quesada *et al.* (2021a):

$$p(X_{0:T}) = p(X_0) \prod_{t=0}^{T-1} p(X_{t+1}|X_t) \qquad (2.9)$$

This combines the GBN structure with a temporal component where the child nodes of a given time slice can only have parents from 1 time slice prior. However, in some cases, connections between t and observations $> t + 1$ may occur, in which high order Markovian models (known as a high order DGBN, or HO-DGBN) are needed. High-order Markovian models

exponentially increase the complexity of the network, and therefore require significant training time (Quesada *et al.,* 2021a). To reduce model complexity, the arcs can be directed to only the most recent time slice, known as a transition network (Quesada *et al.,* 2021b), as this is often the target of interest when constructing a DBN. An HO-DGBN can be represented as explained by Santos *et al.* (2014), where X_t is $X_t = \{X_t^1, X_t^2, \dots, X_t^N\}$ for N variables in T total time slices, and the conditional independencies for the probability distribution is $p(X_{t+1}, X_{t+2}, \dots, X_{t+T} | X_t)$. From here the network can be restricted so that directed arcs are only calculated from the present time slice to any other time slice (inter-time slice), disregarding arcs within a time slice (intra-time slice) (Figure 2.8). The notation for the present time slice can be regarded as t_0 or t_{+T}, where in the context of $p(X_{t+1}, X_{t+2}, \dots, X_{t+T} | X_t)$, t_T is the present time slice (*i.e.*, if $T = 5$, the present time slice is t_{+5}, and t_0 is 5 time slices prior), or $p(X_t | X_{t-1}, X_{t-2}, \dots, X_{t-T})$, t_0 is the present slice (*i.e.*, if $T = 5$, the present time slice is t_0, and t_{-5} is 5 time slices prior). A DGBN typically utilizes its network learning and optimization which incorporates the temporal domain as seen in Table 2.5. Learning structure types, however, remain the same as described in section 2.4.7.1. Once the networks are learned, inferences can be made in the same way as GBN, as described in table 2.4, however the data is forecasted as a time series, considering the temporal sequence of the events.
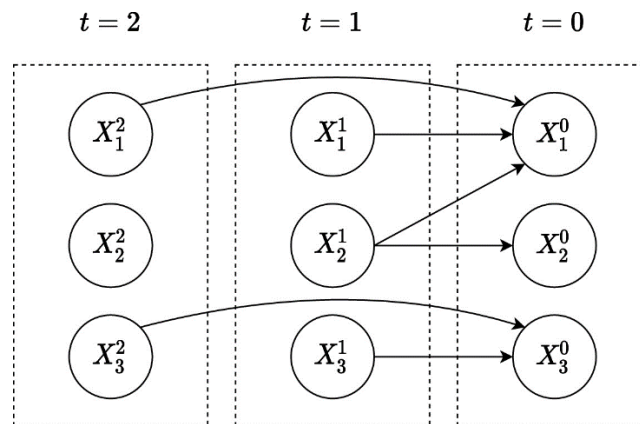
Figure 2.8. Example of an order 2 Markovian network where arcs are only directed toward the latest time-slice. Quesada *et al.,* 2021

Table 2.5. Summary of fundamental dynamic Bayesian network optimization structures as described by Nagarajan *et al.* (2013)

Type	Method	Description	Source
Constraint-Based	Least Absolute Shrinkage and Selection Operator (LASSO)	Estimates coefficients via L_1-constrained least squares where only nonzero coefficients determine significant dependence.	Tibshirani, 1996; Meinshausen & Bühlmann, 2006
	James–Stein Shrinkage	Estimates the covariance matrix by shrinking the correlation coefficients to zero and the variance to their median.	James and Stein, 1961; Opgen-Rhein and Strimmer, 2007
Score-Based	Statistical Inference for Modular Networks (SIMoNe)	Based on LASSO, optimizes the L_1 penalization by searching for latent clustering of the network that returns the lowest BIC.	Chiquet *et al.*, 2009
Heuristic	First-Order Conditional Dependencies Approximation	The DAG is approximated by a first-order conditional dependencies graph.	Lèbre, 2009
	Particle Swarm Optimization (PSO)	Local search space which implements many simultaneous 'particles' to search for the local minimum and updates the velocity after each iteration.	Sahin *et al.*, 2004
	PSO with Order Invariant Encoding	Extension of the PSO for HO-DGBN, where computational costs are reduced by representing the network as vectors of natural numbers with a constant length.	Quesada *et al.*, 2021b
Hybrid	Dynamic Max-Min Hill-Climbing (DMMHC)	Adaptation of MMHC by minimizing the search space and maximizing the network score, for each node and time slice.	Trabelsi, 2013
	Reversible Jump Markov chain Monte	Randomly generates an initial model and for each iteration modifies it which can be either accepted or rejected.	Green, 1995

Type	Method	Description	Source
	Carlo (RJ-MCMC)		

DGBN considers the temporal domain when constructing the interactions between variables as they change over time. Such methods are useful in macroeconomics (Dabrowski *et al.,* 2016; Lytvynenko *et al.,* 2020), genetics sequencing (Perrin *et al.,* 2003; Suter *et al.,* 2022) and even environmental monitoring (Chang *et al.,* 2023; Valero-Leal *et al.,* 2002). Several studies have used static BN and GBN to enquire about the effects of the interaction of algal biomass with various environmental variables (Rigosi *et al.,* 2015; Moe *et al.,* 2016; Shan *et al.,* 2019; Jiang *et al.,* 2021; Jackson-Blake *et al.,* 2022). There are currently no studies applying a DGBN to understand the interaction effects of various atmospheric and lake physical variables on algal biomass dynamics, as this requires significant data (Jiang *et al.,* 2021), however, it is considered a desirable product for future analysis (Shan *et al.,* 2019). This research proposes the use of satellite-derived chl-*a* concentrations to fill the DGBN data requirements, thus incorporating the temporal domain.

2.5 Conclusion

The impacts of climate change on algal biomass dynamics are complex, with differing impacts depending on lake morphometry, watershed environments, algal community compositions, and anthropogenic influences. Limited comprehensive analyses have been performed to identify the interactions between climate change and algal biomass dynamics, mainly due to the lack of in situ observations. Satellite remote sensing provides new avenues for determining interaction effects of climate and algal biomass dynamics by supplementing in situ observations with estimates retrieved from historic image archives suing state-of-the-art algorithms. Remote sensing of algal biomass in freshwaters is difficult, however, as the atmospheric and OAC in waters can introduce errors in retrieval algorithms. New remote sensing data products are currently available, allowing researchers to focus on improving our

understanding of how these variables are changing over time, and how this change coincides and is affected by other environmental data.

Chapter 3: Study Area and Data

3.1 Study Area

Five North American Great Lakes have been selected for this study; Lake Erie, Lake Winnipeg, Lake Athabasca; Great Slave, and Great Bear Lake. These lakes were selected as they are large and therefore resolvable using coarse spatial resolution satellite imagery (1km grid) and coarse climatic gridded data (9km grid) and provide a range of temporal to arctic lake systems with varying climatic and land use conditions. Gridded data within each lake and the surrounding catchment will be used in all subsequent chapters of this thesis. A map of the study area along with the general land use of the regions can be found in Figure 3.1.
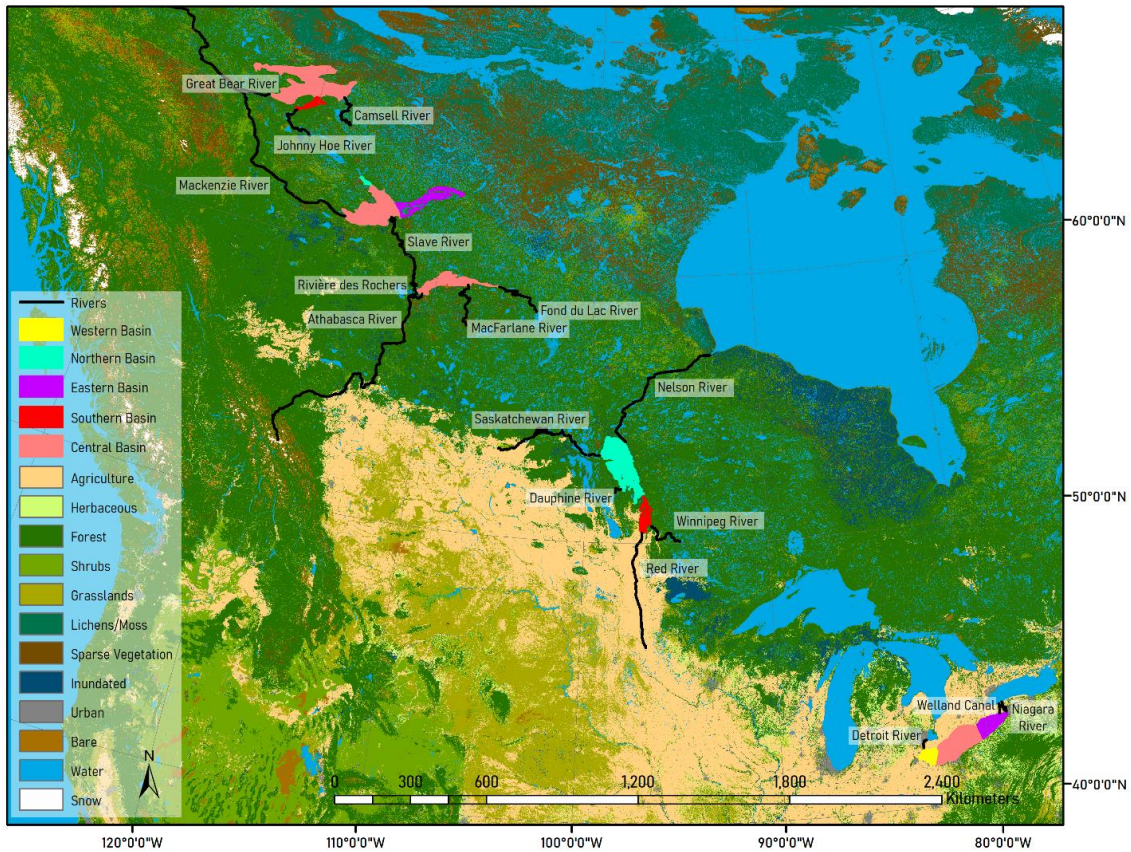
Figure 3.1. Map of study lakes and their corresponding basins and rivers. Land cover delineation provided by ESA CCI Landcover (http://maps.elie.ucl.ac.be/CCI/viewer/download.php).

3.1.1 Lake Erie

Lake Erie is the smallest by volume of water of the Laurentian Great Lakes, with a surface area of 25,744 km^2, and is relatively shallow, with an average depth of ~19 m and composed of three distinct basins: Western, Central and Eastern (EPA, 2020). The Western basin of Lake Erie is the shallowest, with an average depth of ~7.4m, and the most turbid. The central basin is moderately deep, with an average uniform depth of ~18.3 m, while the Eastern basin is the deepest, with an average depth of 24 m (EPA, 2020). The lake has a retention time of ~2.6 years, and 80% of the lake's inflow is a result of the Detroit River. Approximately 12 million people live within the Lake Erie watershed, in which a significant amount of toxic ABs have occurred, prompting significant research by various agencies. Most of the watershed is composed of urban and agricultural regions, which have contributed greatly to the observed eutrophication. The lake is dimictic, in which mixing and subsequent internal loading occur during spring and fall, with summer and winter stratification. As Lake Erie is relatively shallow, most years result in high ice cover; however, frequent breaks and melts occur (Fujisaki *et al.,* 2012).

3.1.2 Lake Winnipeg

Lake Winnipeg is the 10th largest in the world, with a surface area of 23,750 km^2, is very shallow, with an average depth of ~12 m, and composed of two basins: north and south (Government of Manitoba, 2021). The north basin is the largest and deeper of the two basins, with an average depth of ~13.3m compared to ~9m of the south basin, while the south basin remains more eutrophic. The lake has a retention time of 3-5 years, with four major tributaries: Red, Winnipeg, Saskatchewan, and Dauphine rivers (Government of Manitoba, 2021). Approximately 7 million people live within the Lake Winnipeg watershed, in which the population of toxin-producing cyanobacteria has increased greatly since the 1990s, due to an increase in nutrient loading (Government of Manitoba, 2021; Binding *et al.,* 2018). The lake is polymictic, in which mixing events occur at irregular intervals, often triggered by wind and precipitation events due to the shallowness of the lake. The lake experiences annual ice coverage with some breakup throughout the winter months. As the south basin is shallower

than the north, it experiences an ice-off date ~2 weeks earlier than the north basin (ECCC, 2020).

3.1.3 Lake Athabasca

Lake Athabasca is the smallest of those selected in this study with a surface area of 7,770 km^2 and is relatively shallow with an average depth of ~20.0 m (Mitchell & Prepas, 1990). There are not well-defined basins within the lake itself; however, the Western side of the lake is significantly more shallow (maximum depth = 16m) than the eastern side of the lake (maximum depth = 124m) (Mitchell & Prepas, 1990). The Athabasca River (merged with the Peace and Birch Rivers) contributes to 53% of the lake's inflow, while Fond du Lac, McFarlne and Birch rivers contribute 21%, 6% and 3% respectively (Mitchell & Prepas, 1990; Leconte *et al.,* 2008). The terrestrial region of the watershed is relatively undeveloped, with the largest human settlement, Fort Chipewyan, located on the Western edge of Lake Athabasca with a population of 852 (Statistics Canada, 2016). The Western edge of Lake Athabasca is also the most turbid due primarily to the inflow of the Athabasca River; however, the majority of the lake is considered oligotrophic with low algal biomass (Mitchell & Prepas, 1990). Indigenous people have reported an increase in algal growth that has been attributed to discharge from oil sands mining operations transported through the Athabasca River (Bill *et al.,* 1994; Timoney, 2008; Parlee & D'Souza, 2019). Lake Athabasca is dimictic, in which mixing occurs biannually (spring and autumn), with winter and summer stratification (Leconte *et al.,* 2008). Of the Northern Great Lakes (Lake Athabasca, Great Bear Lake and Great Slave Lake), Lake Athabasca has the shortest ice cover duration, with an ~16-day shorter ice cover season compared to Great Slave Lake, where midwinter breakup is very uncommon (Bussières, 2008).

3.1.4 Great Slave Lake

Great Slave Lake has a surface area of 28,450 km^2 and is one of the deepest lakes in the world with an average depth of ~88 m (Rouse *et al.,* 2008). The lake is distinguished by three main basins; Central, Northern, and Eastern, in which the Eastern basin is considered the deepest portion and the Northern basin the shallowest (Avalon Rare Metals Inc, 2011;

Schertzer *et al.,* 2008). Despite the depth and volume of the lake, it has a retention time of 16 years with ~75% of the inflow from the Slave River and most of the outflow in the Mackenzie River (Evans, 2000; Gibson *et al.,* 2006). The watershed surrounding Great Slave Lake is relatively undisturbed, with some built-up regions including Yellowknife, Fort Providence, and Fort Resolution. Although oligotrophic, Great Slave Lake has been observed to experience cyanobacterial blooms in Yellowknife Bay in September 2013 (Pick, 2016; Rühland *et al*., 2023). The lake is dimictic and, therefore, experiences two annual mixing cycles (Rouse *et al.,* 2008). Great Slave Lake experiences annual ice cover and has a 10-day later ice on date and a 21-day earlier ice off date compared to Great Bear Lake (Woo & Rouse, 2008).

3.1.5 Great Bear Lake

Great Bear Lake is the ninth largest in the world, with a surface area of 31,153 km^2 and an average depth of ~76 m (Johnson, 1975; Rouse *et al.,* 2008). The lake is distinguished by five arms; McVicar, Keith, Smith, Dease, and McTavish, where McTavish is the deepest and McVicar is the shallowest and considered its own basin (Johnson, 1975; Chavarie *et al.,* 2015). The two main sources of inflow in Great Bear Lake are the Johnny Hoe River (McVicar arm), and the Camsell River (McTavish arm) with the Whitefish River (Smith arm) and the Dease River (Dease arm) also contributing, while the Great Bear River is the main outflow (Keith arm) (Chavarie *et al.,* 2015). Due to the volume and low outflow, Great Bear Lake has the longest residence time of the selected lakes of 124 years (Johnson, 1975). The watershed surrounding Great Bear Lake is relatively undisturbed, with the settlement of Délı̨nę as the largest populated area and some impacts of mining operation (Schindler, 2001). Great Bear Lake is considered an oligotrophic system and, although algal blooms have not been reported, studies have found that productivity has increased from 2003-2018 (Sayers *et al.,* 2020). The lake has traditionally been considered a cold monomictic system; however, recent studies have shown that some regions of the lake are transitioning/currently dimictic (Rouse *et al.,* 2008). Great Bear Lake experiences annual ice cover with the longest ice cover duration of the selected lakes (Woo & Rouse, 2007).

3.2 Data Sources

To ensure a complete time series of lake physical and biological parameters, satellite derived daily data provided by the European Space Agency (ESA) Climate Change Initiative (CCI) Lakes data product (Carrea *et al.,* 2022; Carrea *et al.*, 2023), and European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-Land data as part of the ERA5 climate reanalysis product (Muñoz, 2019) is to be used (2002 – 2020). Of the ESA CCI Lakes data products, Lake Water Leaving Reflectance (LWLR; satellite derived chl-*a* product), Lake Ice Cover (LIC), and Lake Surface Water Temperature (LSWT) is to be used. The ESA CCI lakes product also includes Lake Water Level, Lake Water Extent, and Lake Ice Thickness, but not used herein. Of the ERA5-Land data, 2-m air temperature, Hourly Land Total Precipitation, Hourly Land Surface Net Solar Radiation, Hourly Land Surface/Subsurface Runoff, and Hourly Land 10m U & V Wind Components is to be used.

3.2.1 European Space Agency (ESA): Climate Change Initiative (CCI) Lakes+

The ESA CCI Lakes+ product aims to provide a consistent and valid global lake database with the longest period of combined satellite observations, covering multiple lake wide variables (Carrea et al., 2022). The specific objectives of the project assess climate research community needs, develop algorithms for produce and provide various lake products, validate through independent climate research groups, and generate new interest in earth observation data for inland water research. This research looks to provide validation and application of these datasets for monitoring algal biomass dynamics as they relate to atmospheric and lake physical variables, all of which are impacted by climate change. Each of the data products were developed using different satellites at varying spatial and spectral resolutions. All products have been harmonized to a common $1/120^{o}$ latitude-longitude (ca. 1 km) grid (Carrea *et al.*, 2022) at a daily scale. While the ESA CCI Lakes+ product also provides lake water level (LWL), and lake water extent (LWE) variables, only LWLR, LIC and LSWT were used for the purpose of this thesis, as LWL and LWE do not vary per pixel.

3.2.1.1 Lake Water Leaving Reflectance (LWLR)

The LWLR product is developed using MERIS and Sentinel-3A OLCI satellites, with MODIS algorithms currently in development to bridge the data availability gap between the MERIS and Sentinel-3A launch and decommission dates. The v2.0 product provides normalized water-leaving reflectances (Rw) for visible-NIR radiometric bands corresponding to MERIS/Sentinel-3A bands (443-900 nm) and Aqua-MODIS bands (405-877 nm; 2012-2015). In situ chl-*a* measurements were provided by the Lake Bio-optical Measurement and Matchup Data for Remote Sensing (LIMNADES) database which is maintained by the University of Stirling and the GloboLakes project (https://limnades.stir.ac.uk). The images were pre-processed in the ESA Sentinel Application Platform (SNAP) to identify water extent and dynamic pixel classification using an ANN to determine optical water types (OWTs). Atmospherically corrected images (POLYMER v4.12) were matched with in situ observations within a temporal window of several days (Carrea *et al.,* 2022). The LWLR product development team has tested multiple empirical retrieval algorithms (utilizing the Red-NIR edge bands) and semi-analytical models, which varied in performance depending on OWT (Carrea *et al.,* 2022). The poorer performance of the ANN may have been attributed to the limited number of training and matchup data. Uncertainty mapping is determined via the weighted OWT membership score, along with uncertainty functions from the algorithm validations. Algorithms were validated using a Round-robin comparison, where each algorithm was tested for each OWT (Simis et al., 2023). Data may be missing due to cloud/ice cover, and due to extreme solar zenith angles for high-latitude lakes during some winter months.

3.2.1.2 Lake Ice Cover (LIC)

The LIC product was developed using MODIS Terra/Aqua level 1B top-of-atmosphere (TOA) calibrated radiance products (MOD02/MYD02). Lake water pixels were identified using ESA CCI land cover (v4.0) data at a 150m resolution. The v2.0 and beyond utilizes a random forest based classification for distinguishing ice, cloud and open water classes. The classification algorithm makes use of bands 1-7, solar zenith angle, geolocation (latitude and longitude), and rasterized QA bands (MOD02). The LIC product provides 3 output bands

defining the LIC flags (forms ice/does not form ice), LIC class (1 = water, 2 = ice, 3 = cloud, 4 = bad), and the classification uncertainty, which is defined using the model accuracy assessment via independent statistical validation (Carrea *et al.,* 2022). This data product enables the identification of total ice cover extent, ice-on, and ice-off timing. Some misclassifications may occur due to shoreline mixed pixilation along with some temporal continuity issues with the MODIS Terra and Aqua sensors (Carrea *et al.,* 2022).

3.2.1.3 Lake Surface Water Temperature (LSWT)

The LSWT was developed using a geophysical inversion via an optimal estimation technique. This method utilizes a classification system to identify and therefore exclude all non-water pixels via the visible-SWIR bands of the ASTR and AVHRR satellites. A Bayesian inference of the LSWT is defined by the difference of the modelled and satellite-retrieved values. The modelled data are derived from numerical weather predictions as inputs of a radiative transfer equation (Simis *et al.,* 2020). Multiple differing satellites are used to compare modelled and observed values, where inter-sensor adjustments are made using AVHRR as a reference. Uncertainty is calculated defined by the standard deviation of the estimated error distribution (Carrera *et al*., 2023) which are used with water detection scores, sensitivity of the retrieval and the chi-square value. Quality indicators are defined using the uncertainty measure, distance from shoreline, chi-square test etc. A quality value of 0 indicates no data, 1-3 indicates poor data quality, while 4-5 indicates acceptable quality. Missing daily data prior to 2006 is due to the lower coverage of ATSR at this time (Simis *et al.,* 2020).

3.2.2 ERA5-Land Hourly Data

ERA5-Land is a reanalysis data product offering a greater spatial resolution (0.1º by 0.1º; 9 km native grid), compared to its ERA5 counterpart (0.5° by 0.5°; 31 km native grid). The reanalysis product uses modelled data in conjunction with earth observations (terrestrial and atmospheric) to provide simulated hourly global climate data from 1950 to present. Factors such as air temperature, humidity and pressure are also corrected for altitude. As the ERA5 data are simulated, uncertainty occurs, with uncertainty increasing with the decrease in time

(early years have higher uncertainty than present). ERA5-Land is widely used for many surface modelling applications, informing policy, business and decision making. For the purposes of this thesis, several ERA5-Land products are used to inform multiple lake physical and atmospheric impacts on algal biomass dynamics: Total Precipitation (PPT), Surface Net Solar Radiation (SNSR), Surface Runoff (SR), Subsurface Runoff (SSR), Wind Speed (WS), Lake Mixing Level Depth (LMLD) and Air Temperature (T2m).

3.2.2.1 Total Precipitation (PPT)

The ERA5 hourly data product models the total precipitation (snow and rain) as modelled by the sum of large-scale and convective precipitation patterns. The large-scale precipitation data is derived from the ECMWF integrated forecasting system (IFS) cloud scheme, which represents the formation of clouds by changes in atmospheric conditions (pressure, temperature, and moisture). The convective precipitation is generated by the IFS convection scheme, which identifies convection at smaller scales. All modelled precipitation data are measured in meters (m) and does not factor fog, dew, or atmospherically evaporated water (Muñoz, 2019). Precipitation is an important component in the mixing of the water column and mobilization of surface and subsurface nutrients.

3.2.2.2 Surface Net Solar Radiation (SNSR)

The ERA5 hourly data product models the Surface Net Solar Radiation (SNSR), which measures the amount of direct and diffuse shortwaves radiation (0.2-0.4 µm) at the earth's surface and subtracted by the surface albedo. Radiative flux is measured at the top-of-atmosphere (TOA) and takes into account both the absorption and scattering of radiance within the atmosphere due to clouds, aerosols, and atmospheric gases. All modelled SNSR data are measured in joules per square meter (J m^{-2}) (Muñoz, 2019). Solar radiation is an important component of algal growth and influences water temperature/stratification.

3.2.2.3 Surface Runoff (SR)

The ERA5 hourly data product models surface runoff as the water flux that exceeds the infiltration rate using both the Tiled ECMWF Scheme for Surface Exchanges over Land

(TESSEL) and the Hydrology TESSEL (HTESSEL) models. Runoff is calculated for up to six surface classes (bare, low/high vegetation, intercepted water, and shaded/exposed snow) and two water classes (open/frozen water), with four vertically stacked soil classes, whose depths are approximated by geometric relation. The model includes a soil heat budget (using Fourier diffusion law) and takes into account soil water melt/freeze up, where the top layer is a net ground heat flux and the bottom a zero-flux. The model determines the point of saturation, whereby any excess water (via rainfall) is mobilized either through infiltration or surface runoff (Balsamo *et al.,* 2009). Vegetation is characterized using the Biosphere-atmosphere Transfer Scheme (BATS) model, which classifies land surface biomes and provides vegetation characteristics (high or low) and area fraction (high and low) (Balsamo *et al.,* 2009; ECMWF, 2016). Surface runoff is measured in meters (m) and represents the mean over a grid square. This product does not indicate mobilization or directionality of the runoff. Surface runoff is an important component of how much terrestrial nutrient export within the watershed.

3.2.2.4 Subsurface Runoff (SSR)

The ERA5 hourly data product models the subsurface runoff as the portion of water that infiltrates the soil column using the TESSEL and HTESSEL models. The model utilizes Darcy's law to calculate subsurface water flux and calculates the soil water equation via the four-layer soil discretization plus the heat budget equation (Balsamo *et al.,* 2009). The top boundary of the subsurface runoff is the infiltration rate/surface evaporation, while the bottom boundary assumes free drainage, and all layers in-between exhibit water sinks via root extraction over vegetated regions (Balsamo *et al.,* 2009). All subsequent equations are processes are calculated in the same way as surface runoff as described in section 1.5.6. Subsurface runoff is measured in meters (m) and represents the mean over a grid square. This product does not indicate mobilization or directionality of the runoff. Subsurface runoff is an important component that determines the amount freshwater export and subsurface nutrient mobilization.

3.2.2.5 10m U & V Wind Components (WS)

The ERA5 hourly data product models both the Eastward (U) and Northward (V) 'neutral' wind speeds. Wind speeds models incorporate surface roughness for inhomogeneous terrain through the introduction of 'exposure correction'. This correction metric makes use of the blending height and roughness length for varying surface conditions (low and high vegetation, buildings, etc.) (ECMWF, 2016). Neutral wind speeds assume the air is stable and stratified, thereby providing slower speeds than real conditions during stable air masses and higher than real conditions during unstable air masses. Wind speeds are measured in meters per second (m s^{-1}) at 10 m above the earth's surface. Wind speed is an important metric in the lateral mobilization of algae and promotes lake mixing in shallow waters.

3.2.2.6 Lake Mix-Layer Depth (LMLD)

The ERA5 hourly data product models the depth at which freshwater bodies have a near uniform temperature gradient with depth (from the surface) and is therefore well mixed. Therefore, the model indicates two layers, the upper is the mixing layer and the lower is the thermocline (Muñoz, 2019). The mixing layer depth is modelled through the use of the FLake model, which is done through the analysis of the lake temperature profile, wind speed, humidity, precipitation, and both long- and shortwave radiation (Betts *et al.,* 2020). The LMLD is measured in meters (m) and is important in determining the duration of lake mixing events and subsequent redox potential of benthic nutrients.

3.2.2.7 2m Air Temperatures (T2m)

The ERA5 2-m air temperature (T2m) represents the temperature (K) of the air as it is 2 m above the earth's surface. This product takes into account various atmospheric conditions and is calculated by interpolating the Earth's surface with the lowest level of the H-TESSEL model (ECMWF, 2016).

3.3 Conclusion

Algal biomass dynamics remain difficult to identify due to the sparsity of data with traditional in situ methods, water sampling expenses, and the remote location/quantity of lakes. The ESA CCI Lakes+ data provides daily surface water chl-a concentrations for thousands of lakes at a 1km grid, along with LSWT and LIC. This data is utilized in the following chapters to explore algal biomass dynamics as they relate to lake physical and atmospheric forcings change over time.

Chapter 4 utilizes the LWLR product to calculate seasonal and annual mean/max chl-a concentrations, peak chl-a timings, and establish lake specific algal biomass anomalies. The LSWT product was used to calculate seasonal and annual mean/max LSWTs, peak LSWT timings, and LSWT anomalies. Finally, The LIC product can be used to identify lake ice on/off timings and lake ice duration. Such derived LSWT and LIC data from the ESA project can be used to determine the role LIC plays in the observed seasonal and annual algal biomass trends and variability over the past 19 years. LIC is an important factor for ice forming lakes and their phenology, potentially affecting algal community structures, the timing of nutrient availability, stratification, mixing timings, and DO (Rumyantseva *et al*., 2019). LSWT is widely regarded as on of the most important variables in algal growth, directly impacting productivity (Paerl & Otten, 2013; Borowitzka *et al.,* 2016; Gobler, 2020).

The ESA CCI data provides complete LWLR, LSWT and LIC data for all lakes as established in section 3.1, providing an opportunity for identifying the trends and variability for each lake and their subsequent basins. These data can be further used to determine if LIC is an important predictor for annual and seasonal algal biomasses, as seen in chapter 5, using various machine learning techniques as seen in section 2.4.5.

The ERA5 data allows for an expansion of analysis beyond LIC, to other lake physical and atmospheric variables, all of which are impacted by climate change, as explored in chapter 6. Such products provide daily observations of all variables, allowing for comprehensive network analysis to determine how these variables interact while changing over time using a DGBN as seen in section 2.4.7.3. The variables used impact algae in different ways: T2m and

SNSR impacts algal productivity (Paerl & Huisman, 2008), PPT, SR and SSR act as vectors for external nutrient mobilization and lake flushing (Coffey *et al.,* 2019), LMLD indicates the level of mixing occurring within the lake water column, affecting nutrient availability and community compositions (Salmaso, 2010), while WS can act as a disruptor of surface algae and can initiate mixing in shallower waters (Huang *et al* 2016; Zhou *et al* 2021). The temporal resolution of the ESA CCI Lakes+ and the ERA5-Land data allows for an identification of temporal lag effects between these variables and the algal biomass response at as fine as the daily scale.

The impacts of lake physical variables and atmospheric forcings on algal biomass dynamics remain relatively unknown, where remote sensing technologies improve data availability and provide avenues for longer time series observational studies. The following chapters will explore these data products in detail, and identify trends, variances, predictive capacities and interaction effects between the predictor variables and algal biomasses, to better understand their dynamics and how they may change in a changing world.

Chapter 4: Is Lake Ice a Driver of Trends and Variability of Algal Biomass in North American Great Lakes?

4.1 Introduction

It has been hypothesized that the perceived increase in global algal blooms (ABs)/biomass is attributed to climate change (Paerl & Huisman, 2008; Pick, 2016; Winter *et al*., 2011). The primary driver of AB formation in mid-latitude lakes is predominately attributed to anthropogenic eutrophication through intensive agriculture, reduction of natural wetlands/forested environments, increased phosphorus/nitrogen-rich waste waters, and dissemination of invasive species (Mitsch, 2017; Waters *et al*., 2021). Recent studies, however, have indicated that changes in global climate provide favorable conditions for AB formation and biomass growth (Wells *et al*., 2015; Trainer *et al*., 2020), even for remote mid (30°N - 60°N) to high (>60°N) latitude lake systems (Ayala-Borda *et al*., 2021; Favot *et al*., 2019). The observed and projected increase in air and water temperatures (Komatsu *et al*., 2007; O'Reilly *et al*., 2015; Filatov *et al*., 2019; Niedrist *et al*., 2018; Woolway *et al*., 2019) are known to prolong the growing season (Adrian *et al*., 2016), improve algae growth rate and (predominately) increase spatial/temporal range of cyanobacteria (Deng *et al*., 2014; Borowitzka *et al*., 2016; Cross *et al*., 2015; Gobler, 2020). Furthermore, increasing global temperature increases thermal stratification rates (Ficke *et al*., 2007; Woolway *et al*., 2020), which fuels anoxia (Nolan *et al*., 2019) and provides a potential increase in cyanobacterial toxicity (Lürling *et al*., 2017), both of which can lead to subsequent suppression of zooplankton, limiting grazing pressures (Lamb *et al*., 2019; Caron & Hutchins, 2012). Factors such as the observed increasing frequency of short but more intense rainfall events (Min *et al*., 2011; Groisman *et al*., 2004), increase nutrient runoff during the growing season (Dalton *et al*., 2018; Jeppenson *et al*., 2011), and when followed by stratification due to drought/heatwave conditions, have been known to significantly increase the probability of AB formation (Reichwaldt & Ghadouani, 2012). The impact of these climatic parameters on AB formation during the growing season is well documented in a variety of lab (Hennon & Dyhrman, 2020; Griffith & Gobbler, 2020), mesocosm (Trochine *et al*., 2010; Moss *et al*., 2003; 2004) and

59

Chapter 4: Is Lake Ice a Driver of Trends and Variability of Algal Biomass in North American Great Lakes?

4.1 Introduction

It has been hypothesized that the perceived increase in global algal blooms (ABs)/biomass is attributed to climate change (Paerl & Huisman, 2008; Pick, 2016; Winter *et al*., 2011). The primary driver of AB formation in mid-latitude lakes is predominately attributed to anthropogenic eutrophication through intensive agriculture, reduction of natural wetlands/forested environments, increased phosphorus/nitrogen-rich waste waters, and dissemination of invasive species (Mitsch, 2017; Waters *et al*., 2021). Recent studies, however, have indicated that changes in global climate provide favorable conditions for AB formation and biomass growth (Wells *et al*., 2015; Trainer *et al*., 2020), even for remote mid (30°N - 60°N) to high (>60°N) latitude lake systems (Ayala-Borda *et al*., 2021; Favot *et al*., 2019). The observed and projected increase in air and water temperatures (Komatsu *et al*., 2007; O'Reilly *et al*., 2015; Filatov *et al*., 2019; Niedrist *et al*., 2018; Woolway *et al*., 2019) are known to prolong the growing season (Adrian *et al*., 2016), improve algae growth rate and (predominately) increase spatial/temporal range of cyanobacteria (Deng *et al*., 2014; Borowitzka *et al*., 2016; Cross *et al*., 2015; Gobler, 2020). Furthermore, increasing global temperature increases thermal stratification rates (Ficke *et al*., 2007; Woolway *et al*., 2020), which fuels anoxia (Nolan *et al*., 2019) and provides a potential increase in cyanobacterial toxicity (Lürling *et al*., 2017), both of which can lead to subsequent suppression of zooplankton, limiting grazing pressures (Lamb *et al*., 2019; Caron & Hutchins, 2012). Factors such as the observed increasing frequency of short but more intense rainfall events (Min *et al*., 2011; Groisman *et al*., 2004), increase nutrient runoff during the growing season (Dalton *et al*., 2018; Jeppenson *et al*., 2011), and when followed by stratification due to drought/heatwave conditions, have been known to significantly increase the probability of AB formation (Reichwaldt & Ghadouani, 2012). The impact of these climatic parameters on AB formation during the growing season is well documented in a variety of lab (Hennon & Dyhrman, 2020; Griffith & Gobbler, 2020), mesocosm (Trochine *et al*., 2010; Moss *et al*., 2003; 2004) and

59

small-scale (spatial/temporal) observational studies (Vilhena *et al*., 2010; Qin *et al*., 2021). However, the impact of overwintering climate conditions on ABs and algal biomass formation is less understood due to logistical challenges and the assumption of ecological dormancy (Block *et al*., 2019; Hampton *et al*., 2015).

Winter air temperatures have increased at a higher rate compared to any other season, with high latitudes experiencing the highest rate of change (Zhang *et al*., 2019a). As a result, there has been an observed decrease in ice duration for mid-high latitude lakes by 106 days per century (1995 – 2016), six times what was observed in earlier periods (1846 – 1995) (Sharma *et al*., 2021). Ice cover has wide-ranging impacts on lake ecosystems: controlling lake mixing (Pilla & Williamson, 2021), the timing of internal loading (Jewson *et al*., 2009), the duration of the growing season (Adrian *et al*., 2016), and the succession of algal community compositions (De Senerpont Domis *et al*., 2007; Salmaso *et al*., 2018; Blenckner *et al*., 2007). Additionally, snow or ice composition can affect the amount of light available under the ice, resulting in under-ice ABs observed in Lake Baikal (Hampton *et al*., 2015; Kalinowska *et al*., 2019). Despite the control that the presence of ice cover has on many biotic and abiotic factors within a lake, there is a lack of observational studies assessing its importance on algal biomass and AB formation. A significant limiting factor in the observational analysis of lake ice phenology and algal biomass interactions is the lack of in-situ measurements. However, remote sensing technologies can use historical satellite image archives to provide comprehensive time-series analysis of large lakes worldwide. Comprehensive data on ice and algal biomass can provide insight into their trends and variability over time and how they may explain their variance.

This study uses three satellite-derived thematic products generated by the European Space Agency (ESA) Climate Change Initiative (CCI) Lakes project (Carrea *et al*., 2022; Carrea *et al*., 2023): lake water leaving reflectance chlorophyll-*a* (LWLR chl-*a*; proxy of algal biomass), lake ice cover (LIC), and lake surface water temperature (LSWT) for five North American Great Lakes along a latitudinal transect (North to South); Great Bear Lake (GBL), Great Slave Lake (GSL), Lake Athabasca (LA), Lake Winnipeg (LW) and Lake Erie (LE). The

dataset (version 2.0.0; Carrea et al., 2022) is used to construct a 19-year time series (2002-2020) of annual LIC parameters (ice-on (ON), ice-off (OF) and ice duration (DR)), annual and seasonal (spring, summer, and fall) LWLR algal biomass parameters (mean and maximum chl-*a* (MC and XC), mean abnormally high chl-*a* (a proxy for ABs, denoting relative anomaly lake chl-*a* concentrations; BC), abnormally high chl-*a* days (BD) and peak chl-*a* (DOY), and annual and seasonal LSWT parameters (mean and maximum LSWT (MT and XT), LSWT anomaly days (AT) and peak LSWT DOY (PT)) for each lake and their basins (to account for spatial heterogeneity as a result of complex lake morphometry). These satellite-derived parameters will be used to achieve the following objectives: (1) assess short-term trends and variabilities in LIC, LSWT and algal biomass parameters from 2002 to 2020; and (2) assess the role of LIC and LSWT parameters in explaining the variance in algal biomass over time in a multivariate space using vector autoregression (VAR). Due to the control that LIC has on the timing of nutrient availability via internal loading (mobilized by mixing) (Huang *et al*., 2022) and the greater rate of change, it is anticipated that LIC is a significant driver of algal biomass trends for the remote Northern Lakes (NL; GBL, GSL, and LA) (Feiner *et al*., 2022) compared to the more anthropogenically developed Southern lake regions (LW and LE). Furthermore, LIC is expected to have a greater impact on algal biomass trends during spring due to earlier light availability and nutrient release, compared to later in the year when temperature and external loading play a more critical role.

4.2 Study Area

Five North American Great Lakes were selected to determine the trends and variability in the chl-*a* parameters along a latitudinal transect (North to South): Great Bear Lake (GBL), Great Slave Lake (GSL), Lake Athabasca (LA), Lake Winnipeg (LW), and Lake Erie (LE). Table 4.1 provides a summary of the general characteristics of the lakes.

Table 4.1. Summary of the study lakes morphometry and their sources. Mean depth encompasses the entire lake (including all basins), residence time indicates the duration in which water remains within the lake, major inflow and outflow rivers

represent the sources where the majority of water enters and exits the lake, and basins indicate regions within a lake where there is minimal hydrological connectivity to other regions within the same lake (often a result of lake morphometry).

Lake	Mean Depth (m)	Surface Area (km^2)	Residence Time (years)	Main Inflow River(s)	Main Outflow River(s)	Basins
Great Bear Lake	~76[1,2]	31,153[1,2]	~124[1]	▪ Johnny Hoe River[3] ▪ Camsell River[3]	▪ Great Bear River[1]	▪ Central ▪ South (McVicar Arm)[1,3]
Great Slave Lake	~88[2]	28,450[2]	~16[4,5]	▪ Slave River[4,5]	▪ Mackenzie River[4,5]	▪ North[6,7] ▪ Central[6,7] ▪ East[6,7]
Lake Athabasca	~20[8]	7,770[8]	~6*	▪ Athabasca River[9] ▪ Fond du Lac River[9] ▪ MacFarlane River[9]	▪ Rivière des Rochers[9]	▪ NA
Lake Winnipeg	~12[10]	23,750[10]	~3-5[10]	▪ Winnipeg River[10] ▪ Saskatchewan River[10] ▪ Red River[10] ▪ Dauphine River[10]	▪ Nelson River[11]	▪ North[10] ▪ South[10]
Lake Erie	~19[12]	25,744[12]	~2.6[12]	▪ Detroit River[12]	▪ Niagara River[13] ▪ Welland Canal[13]	▪ West[12] ▪ Central[12] ▪ East[12]

(1) Johnson, 1975; (2) Rouse *et al*., 2008; (3) Chavarie *et al*., 2015; (4) Evans, 2000; (5) Gibson *et al*., 2006; (6) Avalon Rare Metals Inc, 2011; (7) Shertzer *et al*., 2008; (8) Mitchell & Prepas, 1990; (9) Rasouli *et al*., 2013 (10) Government of Manitoba, 2021; (11) Warrack *et al*., 2017; (12) EPA, 2020; (13) Docker er al., 2021.
*Calculated by dividing the total volume of Lake Athabasca by total inflow (Bennett *et al*., 1973)

The surrounding watersheds of the Northern Lakes (NL; GBL, GSL and LA) are relatively undisturbed, with the settlement of Délı̨nę as the largest populated area near GBL; Yellowknife, Fort Providence, and Fort Resolution near GSL; and Fort Chipewyan near LA (Statistics Canada, 2016; Schindler, 2001). GSL was impacted by the construction of Bennett dam in 1967, affecting seasonal flows, sedimentation, and winter nutrient export (Rühland *et al*., 2023). However, all NLs are impacted by mining operations, such as radium and uranium mining in the GBL watershed (Muir *et al*., 2012), gold mining in GSL (Chételat *et al*., 2019) and LA receiving oilsands runoff via the Athabasca River (Bill *et al*., 1994; Schindler, 2001; Timoney, 2008; Parlee & D'Souza, 2019). Of the NLs, GSL and LA exhibit dimictic mixing cycles (Leconte *et al*., 2008; Rouse *et al*., 2008), while GBL is traditionally cold monomictic;

however, it is anticipated that some regions of the lake are transitioning/currently dimictic (Rouse *et al*., 2008; Rao *et al*., 2011). The NLs are typically oligotrophic, with GSL experiencing some cyanobacterial ABs in Yellowknife Bay, and indigenous groups have reported increased algal growth in LA (Pick, 2016; Parlee & D'Souza, 2019).

The surrounding watersheds of the Southern Lakes (SL; LW and LE) experience significant anthropogenic development, including agriculture, urbanization, and industrial development (Government of Manitoba, 2021; EPA, 2020) (Figure 4.1). Of the SLs, LW and the west basin of LE exhibit a polymictic mixing cycle, where mixing is triggered by wind/precipitation events, while the Central and Eastern basins of LE exhibit a dimictic mixing cycle (Nürnberg & LaZerte, 2016; Karatayev *et al*., 2021). Both LW and LE have experienced significant ABs over the past 20 years, with the majority occurring in the north basin of LW and the eutrophic west basin of LE (Budd *et al*., 2001; LWIC, 2005; Wassenaar *et al*., 2012; Binding *et al*., 2018; Binding *et al*., 2019). Of the selected lakes, only LE experiences intermittent ice cover (Cai *et al*., 2022), with frequent breaks during the ice season, while GBL, GSL, LA, and LW all experience a complete ice cover annually (Howell *et al*., 2009; Ziyad *et al*., 2020).
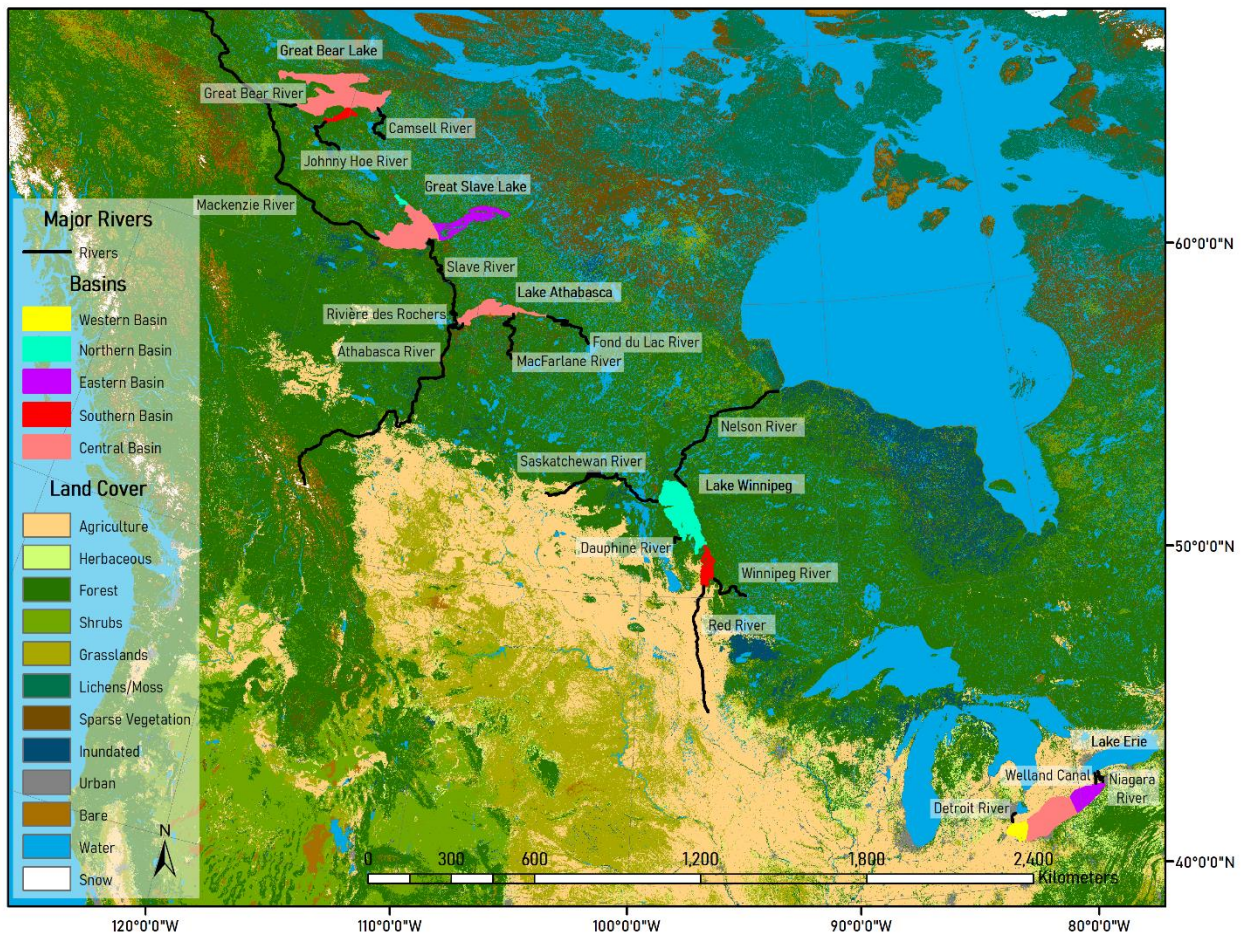
Figure 4.1. Map of study lakes and their corresponding basins and rivers. Land cover delineation provided by ESA CCI Landcover (http://maps.elie.ucl.ac.be/CCI/viewer/download.php).

4.3 Data Sources

Complete annual and seasonal time series (2002-2020) were prepared using a multivariate daily satellite data product available from the European Space Agency's (ESA) Climate Change Initiative (CCI) Lakes project (Carrea *et al*., 2022). The ESA CCI Lakes product is a harmonized dataset that includes various daily metrics for over 2000 of the largest lakes globally georeferenced on a 1/120 degree latitude-longitude (ca. 1 km) grid (Carrea *et al*., 2023). For the purposes of this study, Lake Water Leaving Reflectance (LWLR; satellite-derived chl-*a* product), Lake Ice Cover (LIC), and Lake Surface Water Temperature (LSWT)

are used. Version 2.0.0 of the LWLR thematic product provides chl-*a* (mg m^{-3}) retrievals based on MERIS (2002-2011) and Sentinel-3A OLCI (2016-2020) satellite data, with MODIS (2012-2015) algorithms bridging the data availability gap.

The LWLR product makes use of the Lake Bio-optical Measurement and Matchup Data for Remote Sensing (LIMNADES) database which is maintained by the University of Stirling and the GloboLakes project (https://limnades.stir.ac.uk). The in-situ observations were used to construct empirical, bio-optical and ANN retrieval algorithms combined with dynamic pixel classification of varying optical water types (Carrea *et al*., 2022). The LWLR product does present some limitations, such as the impact of turbid water conditions increasing retrieved algal biomass and potential bias between sensors. The LIC thematic product provides lake cover classes (open water, ice, cloud and 'bad' (unusable) data). It is generated using MODIS Terra/Aqua data processed using a random forest (RF) algorithm (Wu *et al*., 2021). High overall classification accuracy (>95%) has been achieved with the RF algorithm in both spatial (global set of lakes) and temporal (several ice seasons) transferability assessments (Wu et al., 2021). However, it is important to note that the presence of clouds as well as extensive cloud cover periods and low solar illumination angles (> 85°), particularly during the fall freeze-up at high latitudes, introduce classification errors and limit the retrieval of open water and ice cover for several days of the year. (Carrera *et al*., 2022). The LSWT product provides surface water temperature measurements (Kelvin) derived from MODIS, Sentinel-3A/B SLSTR and ERA5 data utilizing a geophysical inversion technique with various modelled priors (Carrea *et al*., 2023). The limitations of the LSWT product include issues with water pixel identification, static emissivity assumptions, and sub-optimal lake-satellite geometries. To identify seasonal threshold timings for the study period (2002-2020), the ERA5-Land hourly 2-m air temperature (Kelvin) product provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) was used (Muñoz, 2019). Given the known limitations of the current version of the CCI Lakes dataset, the data was preprocessed to mitigate outliers and fill missing data gaps, such as eliminating pixels with high uncertainty, and interpolating missing values temporally.

4.4 Methods

To identify features such as peak chl-*a* concentrations and ON/OF timings, missing data must be filled, and therefore preprocessing was necessary to construct a complete time series (2002-2020) of LIC, LSWT, and algal biomass. R software version 4.3.1 was used to develop five algal biomass parameters (mean chl-a (MC), max chl-a (XC), mean abnormal chl-a (BC), abnormal chl-a days (BD) and peak chl-a DOY (PC)), three LIC parameters (ice on (ON), ice off (OF), and ice duration (DR)), and four LSWT parameters (mean LSWT (MT), max LSWT (XT), anomaly LSWT (AT) and peak LSWT DOY (PT)) . Daily rasters were stacked via water years from 1 September 2001 to 31 August 2020 and masked to daily pixel quality flags (LIC data product flags pixels as "bad" or "cloud", LSWT product flags uncertainty by the "quality_level" band, where a pixel value of 1 is not recommended for use, 2-3 are for use at the user's discretion, and 4-5 are recommended for use), which converted the corresponding pixel values to "NA". Each stacked pixel was used to represent a time series of 6,940 days, where missing data values (flagged as NA) were temporally interpolated linearly. Annual LIC is determined through water years (1 September to 31 August) to cover the start (ON) and end (OF) of each ice season. Algal biomass (chl-*a* concentrations) and LSWT were calculated annually by Julian calendar years to cover the beginning (post OF) and end (pre ON) of the growing season.

Seasonal data were aggregated within three seasons: spring, summer, and fall. Seasonal date thresholds differ between lakes due to significant latitudinal and geographic distances. To determine seasonal thresholds, the hourly ERA5 2-m air temperature is used to calculate a daily mean, which was then used to calculate the spatial mean within the lake boundary. The daily spatial 2-m air temperature means were then used to calculate the DOY mean for the study period (2002-2020). Mean DOY temperatures ≤ 273.15 K were manually classified as the winter season and excluded from any analysis of LSWT and algal biomass. The remaining days were entered into an unsupervised univariate k-means classifier to determine the threshold between two clusters. Cluster 1 represents two classes; spring and fall, where spring is defined

as the dates after winter (air temperatures > 273.15 K) but before cluster 2 (summer), and fall is defined as the dates after cluster 2 but before winter.

Five algal biomass parameters were calculated for the interpolated daily algal biomass rasters: MC, XC, BC, BD, and PC. Since ABs are unlikely to occur in NLs, abnormal chl-*a* is used as a proxy. The interquartile range (IQR) and third quartile values (Q3) were calculated using a natural log transform of the chl-*a* to eliminate outliers due to potential retrieval error (*e.g.*, chl-*a* concentrations exceeding 500μg L^{-1} in GBL). Retrieval error may be a result of mixed pixelation, turbid water conditions, LIC presence, and sun glint. The following equation was used to establish the threshold for determining outliers due to retrieval error: $O = Q3 + (1.5 * IQR)$ where O is the outlier threshold. Natural log scaled chl-*a* concentration pixels that exceed this threshold were classified as an outlier and reclassified as "NA". The annual/seasonal MC and XC were calculated per pixel. The PC represents the DOY where chl-*a* was highest for each pixel. To determine BC, a threshold was calculated using a z-score ($Z = (x - \mu)/\sigma$), where the mean (μ) and σ were calculated from the entire spatial extent of the lake over the entire study period (2002 – 2020). The chl-*a* concentrations were also scaled with a natural log to better represent a normal distribution. Any pixel with a z-score of >2 was classified as an abnormal chl-*a* concentration (BC). Annual/seasonal BC represents the average of only abnormal values for that given lake and year/season at each pixel. If no chl-*a* concentration exceeds a z-score of 2 at a given pixel position at any point in the year/season, it is classified as "NA" and excluded from spatial mean calculations, unlike XC, which identify max chl-*a* concentrations for every pixel. This means that we are only looking at pixels/regions of the lake where chl-*a* concentrations are in the highest of the distribution, typically referring to the most productive waters of each lake. The thresholds are calculated lake wide and not at the basin level. To determine the BD, chl-*a* concentrations that exceeded the BC threshold were reclassified as value of 1, while those below the threshold were reclassified as 0. The annual/seasonal sum was calculated at each pixel to determine the number of days above the threshold.

The LIC data product was used to determine each lake's annual ON, OF, and DR per pixel. The ON/OF and DR were only calculated annually, since these parameters occur only once a year during the winter. To determine the lake ON/OF date, an annual stack of daily LIC rasters was created by water year (1 September to 31 August). The raster was reclassified so that all pixels labelled 3 ("cloud") and 4 ("bad") were converted to a "NA" value. At each pixel, the stacked raster (1:365 or 1:366 per leap year) was linearly interpolated to provide a value between 1 (water) and 2 (ice) where observations were not available. As each raster stack begins on 1 September (before ice formation), the ON date begins earlier in the stack compared to the OF. The ON date was calculated by determining the position within each stack where the ice pixels (values > 1) persisted for at least 14 days. The OF date was calculated with the same process; however, the raster stack was inverted, and the position is corrected by subtracting the length of the raster stack from the retrieved position. The DR was determined by calculating the difference between the OF and the ON dates for each pixel.

A total of four annual/seasonal LSWT parameters were calculated from the interpolated daily LSWT rasters: MT, XT, PT, and AT. Pixels flagged with an uncertainty of ≤ 3 were reclassified as NA values, where the missing observations were then linearly interpolated temporally for each pixel. Mean and maximum annual/seasonal LSWTs were calculated per pixel. Peak LSWT DOY represents the DOY where each pixel's LSWT values were highest. To determine LSWT anomaly days, a LSWT threshold was calculated using the sum of the median ($\tilde{\mu}$) and standard deviation (σ) LSWT across the entire lake/basin from 2002-2020. Any daily pixel that exceeded the threshold was converted to a 1, while those below the threshold were converted to 0. The per pixel sum was then calculated annually/seasonally, where each pixel represented the number of days in which the LSWT exceeded the lake-specific threshold.

Global trends in the increase in algal biomass and frequency and severity of AB have been observed to be not accurate but rather a result of increased surveillance in reporting (Hallegraeff *et al*., 2021). It was, therefore, imperative to determine whether monotonic trends existed within each study lake before any meaningful conclusions could be drawn regarding

the relationship between LIC, LSWT and algal biomass parameters. To identify monotonic trends, each lake was separated by basin and season, where a mean pixel value was used for each year, for a total of 19 values (2002-2020) for each lake/basin/season. A Mann-Kendall test (R package 'Kendall' by Mcleod, 2005) was used to determine if a significant trend was present (p-value < 0.1), while a Theil-Sen regression (R package 'RobustLinearReg' by Hurtado, 2020) was used between each parameter and year to calculate the slope (β) for the direction and magnitude of the trend, and the standard error (s_e) for the variability. The s_e indicates how much variance there is between true and predicted values of the regression, where the higher the s_e, the greater the variability. Heteroskedasticity in linear regression residuals was calculated using the Breusch-Pagan test in r ('lmtest' package), where a p-value < 0.1 indicated a change in the magnitude of variance over time.

Vector autoregression (VAR) was employed to determine the interaction of LIC and LSWT with algal biomass over time. VAR models have been used in the past for AB modelling (Liu *et al*., 2007), allowing for simple and flexible multivariate time series modelling, incorporating past and current observations (Cruz *et al*., 2021). VAR is similar to multiple linear regression (MLR) in that it is defined by the linear relationships between each predictor variable and the response variable, where each predictor is defined by a slope coefficient, with all parameters sharing a common intercept. Additionally, VAR is an extension of a univariate autoregression model, in which a linear model of each variable over time is regressed to one or more prior observations in the time series. VAR adds the vector of time and the lagged time of each predictor variable in multiple regressions. The lag component indicates that a prior observation may be related to the subsequent observations, whereby LIC or LSWT measured at a prior interval may impact future algal biomass parameters, not just those at the same time step. Chl-*a* concentrations observed in a lake will likely affect the prevalence of future chl-*a* concentrations (Bégin *et al*., 2020). Furthermore, legacy effects due to significant changes in climatic conditions have been observed in lakes and rivers, altering thermal regimes and affecting biomass (Kendrick *et al*., 2018; Woolway & Merchant, 2018). A standard VAR is calculated using the following equation:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \tag{4.1}$$

where y_t represents a set of K endogenous variables, A_i is a coefficient $(K \times K)$ matrix (where $i = 1 \dots p$), and u_t is a k-dimensional vector of the t residuals where the current time step depends on the previous time step, plus a white noise error metric (Schlabing *et al*., 2014; Pfaff, 2008). The traditional VAR assumes all input parameters represent a stationary process; however, given the nature of the research question, non-stationary processes were expected due to anthropogenic climate change. Deterministic regressors D_t were included in the VAR, where both the intercept and trend of the time series were added to address issues of non-stationarity, resulting in a modified VAR as given in the following equation:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + D_t + u_t \tag{4.2}$$

This study utilized the R package 'vars' (Pfaff, 2008) for model calculations. A VAR is sensitive to over-parameterization when there is a low number of observations and a high number of parameters (Nicholson *et al*., 2017). Lag tests were performed with all potential combinations of LIC and LSWT parameters for each biomass parameter. The lag test with the highest mean adjusted r^2, and the lowest mean Bayesian Information Criteria (BIC) and standard error across all combinations of parameters was selected for use. A MLR was used to determine if a simplified regression without a lag order provideed a similar or superior explanation of variance. Once the lag order was selected, the model performance and error metrics were used to determine which LIC and LSWT parameters provided the best fit over the lagged time series (Table 4.2).

Table 4.2. Summary of performance and error metrics for VAR models were Adj. r^2 = corrected (adjusted) coefficient of determination, AICc = Akaike information criterion second order, BIC = Bayesian Information Criteria, fpe = Final Prediction Error, and s$_e$ = Standard error.

Parameter	Equation	Variables	Measure	EQ (#)
Adj. r^2	$1 - \left[\dfrac{(1 - r^2) \times (n - 1)}{(n - k - l)} \right]$	n = sample size, r^2 = coefficient of determination, k = number of independent variables	Corrected (adjusted) coefficient of determination (r^2) to determine model accuracy	4.3

Parameter	Equation	Variables	Measure	EQ (#)
AICc	$-2\big(l(\theta)\big) + 2k$ $+\left[\dfrac{2k(k+1)}{(n-k-1)}\right]$	n = sample size, k = number of model parameters, $l(\theta)$ = the log-likelihood	Akaike information criterion (AIC) second order for model quality. Useful for datasets with small sample size	4.4
BIC	$-2 \times l(\theta) + k \times \log(n)$	n = sample size, k = number of model parameters	Bayesian Information Criteria (BIC) for model quality	4.5
fpe	$\left(\dfrac{T + n^*}{T - n^*}\right)^k \det\left(\tilde{\Sigma}_u(n)\right)$	T = the time step, n^*= total number of parameters in each equation, $\tilde{\Sigma}_u$= sum of residuals and white noise of the lag order (n)	Final Prediction Error at the given lag-timestep interval (difference between actual and forecasted values)	4.6
s_e	$\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{df}}$	y_i = the observed value, $\hat{y}_i$ = the predicted value, df = degrees of freedom	Standard error of the residuals indicating model fit and variance	4.7

To assess the relative importance of LIC compared to LSWT parameters in explaining the variance of the change in algal biomass over time, performance metrics across all models were aggregated into three categories for each lake, basin, and season: (1) models where only LIC parameters were included; (2) models where only LSWT parameters were included, and (3) models where at least one LIC and one LSWT parameter was included. To determine which model and, subsequently, which input predictor parameters provided the best fit for each algal biomass parameter, a ranking system was devised using all the performance metrics outlined in Table 4.2. Each model performance metric was ranked for all lakes, basins, and seasons, where the lower the rank value, the greater the comparative performance of the model. The model with the lowest mean rank value across all performance metrics was selected as the best-performing. Individual parameter performance was determined by the mean rank when the parameter was included in the model. Performance metrics were also aggregated into two lake groups; Northern (GBL, GSL and LA) and Southern (LW and LE), to identify differences in parameter performance.

The residuals derived from the VAR models provided insight into the variability of the response of algal biomass to the LIC and LSWT parameters. Residuals showing greater

fluctuation were indicative of increasing variability in the algal biomass parameter, while the s_e quantified the magnitude. All input parameters were rescaled by their mean and standard deviation to improve comparisons between parameters and lakes/basins.

4.5 Results

4.5.1 Seasonal Threshold Identification

The data are aggregated into seasonal measures to compare the annual LIC parameters with the daily algal biomass measurements. Each season was used to build its own time series from 2002-2020. Of the ERA5 mean lake-wide 2-m air temperature (2002-2020), the k-means classifier defined the seasonal (spring, summer, fall) thresholds, as seen in Table 4.3.

Table 4.3. Delineated seasonal thresholds as defined by k-means from ERA5 2-m air temperature for comparison of LIC parameters to algal biomass measurements.

Lake and Season	Start DOY	Start Date	End DOY	End Date
GBL Spring	137	17 May	174	23 June
GBL Summer	175	24 June	258	15 September
GBL Fall	259	16 September	286	13 October
GSL Spring	130	10 May	173	22 June
GSL Summer	174	23 June	259	16 September
GSL Fall	260	17 September	300	27 October
LA Spring	120	30 April	164	13 June
LA Summer	165	14 June	257	14 September
LA Fall	258	15 September	302	29 October
LW Spring	108	18 April	156	5 June
LW Summer	157	6 June	270	27 September
LW Fall	271	28 September	310	6 November
LE Spring	72	13 March	136	16 May
LE Summer	137	17 May	287	14 October
LE Fall	288	15 October	366	1 January

For GBL, the seasonal duration lasts 37, 83 and 27 days for spring, summer, and fall, respectively, with GSL lasting 43, 85, and 40 days, LA lasting 44, 92, and 44 days, LW lasting 48, 113, and 39 days, and LE lasting 64, 150 and 78 days. These seasonal thresholds define the

timeframe where lake/basin-wide means are derived for all parameters for each year of analysis.

4.5.2 Trends of Timeseries

Theil-Sen slopes (Figure 4.2) of each annual and seasonal mean algal biomass parameter (MC, XC, BC, BD and PC), LIC parameter (ON, OF and DR) and LSWT parameter (MT, XT, AT and PT) were used to derive the β, while a Mann-Kendall (MK) test was used to determine significance ($p < 0.1$) (Figures 4.3 for algal biomass time series, 4.4 for LSWT time series and 4.5 for LIC time series). The number of tests is based on the number of lakes and basins (5 lakes, 10 sub-basins), and the number of seasons (3 season, 1 annual). The total number of tests across all 15 lakes/basins and all 4 seasons for each parameter is 60. A total of 300 MK tests were conducted (15 lakes/basins, 4 season, and 5 parameters) for algal biomass parameters, 240 MK tests (15 lakes/basins, 4 season, and 4 parameters) for LSWT parameters, and 45 MK tests (15 lakes/basins, 1 season, and 5 parameters) for LIC parameters.

For MC, XC and BC parameters, across all seasons and lakes basins, 63.33% (114 out of 180 total MK tests) returned significant trends, where 95.61% (109/114 MK tests) were positive (increasing) ($\beta = 0.01 – 0.75$ µg L^{-1} yr^{-1}). The spring season shows the greatest difference in MC, XC and BC trends between NL and SL, where NL returned a mean β of 0.16 µg L^{-1} yr^{-1} compared to -0.05 µg L^{-1} yr^{-1} of SL. While most trends were positive, LW consistently returned a (decreasing) trend for BC, where 16.67% (2/12 MK tests), were significant. For BD and PC parameters, across all seasons and lakes, only 20.00% (24/120 MK tests) returned significant trends, where 45.83% (11/24 MK tests) were positive ($\beta = 0.20 –$ 2.50 days yr^{-1}) and 54.17% (13/24 MK tests) were negative ($\beta = -0.25 – -1.93$ days yr^{-1}). Summertime BD returned a higher trend compared to any other season. The trend of PC varies between lakes and seasons, where PC is occurring earlier in the Fall for NL and later for SL, while the inverse is true for the summer (figure 4.2). Map examples of MC can be found in supplementary figure B.7.1.

73

For ON, OF and DR parameters, across all lakes, only 17.78% (8/45 MK tests) returned significant trends, of which, 100% were negative (decreasing) (-0.88 – -1.08 Days yr^{-1}). The remaining 82.22% of tests across all LIC parameters and lakes were not significant, exhibiting little change, either positive (increasing) or negative (decreasing). Of the non-significant trends, little change was found for ON in NL, however most SL returned a negative (decreasing) trend. Of the non-significant trends, the OF and DR were negative (decreasing) for all lakes with the exception of LW.

For MT and XT parameters, across all seasons and lakes, only 11.67% (14/120 MK tests) returned significant trends, where 85.71% (12/14 MK tests) were positive (0.03 – 0.14 K yr^{-1}). Both summer and fall show an increase in MT and XT across all lakes, with the exception of LW. However, Spring MT is highly variable across all lakes, while Spring XT shows a general increase for NL, and a decrease for SL. Annual MT show a general decreasing trend for NL and an increasing trend for SL, while annual XT is increasing for all lakes. For AT and PT parameters, across all season and lakes, only 9.17% (11/120 MK tests) returned significant trends, where 27.27% (3/11 MK tests) were positive (0.20 – 0.30 Days yr^{-1}) and 73.73% (8/11 MK tests) were negative (-0.30 – -0.67 Days yr^{-1}). For complete results, refer to supplementary tables B.7.1 to B.7.4.
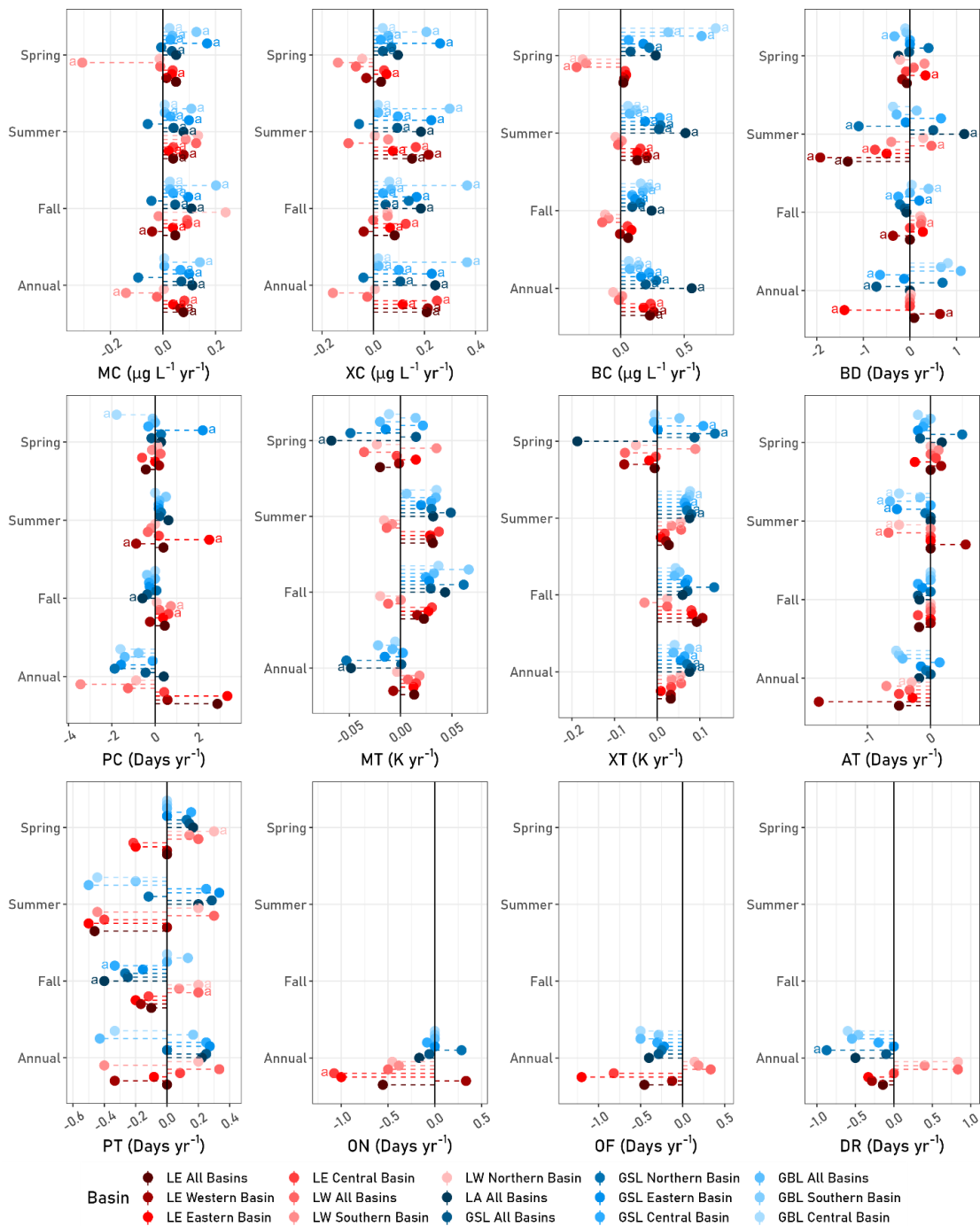
Figure 4.2. Theil-Sen slope calculated for 15 lake/basins at an annual/seasonal scale for 12 parameters. The presence of an "a" indicates a significant (p < 0.1) time-series as calculated by a Mann-Kendall test. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.
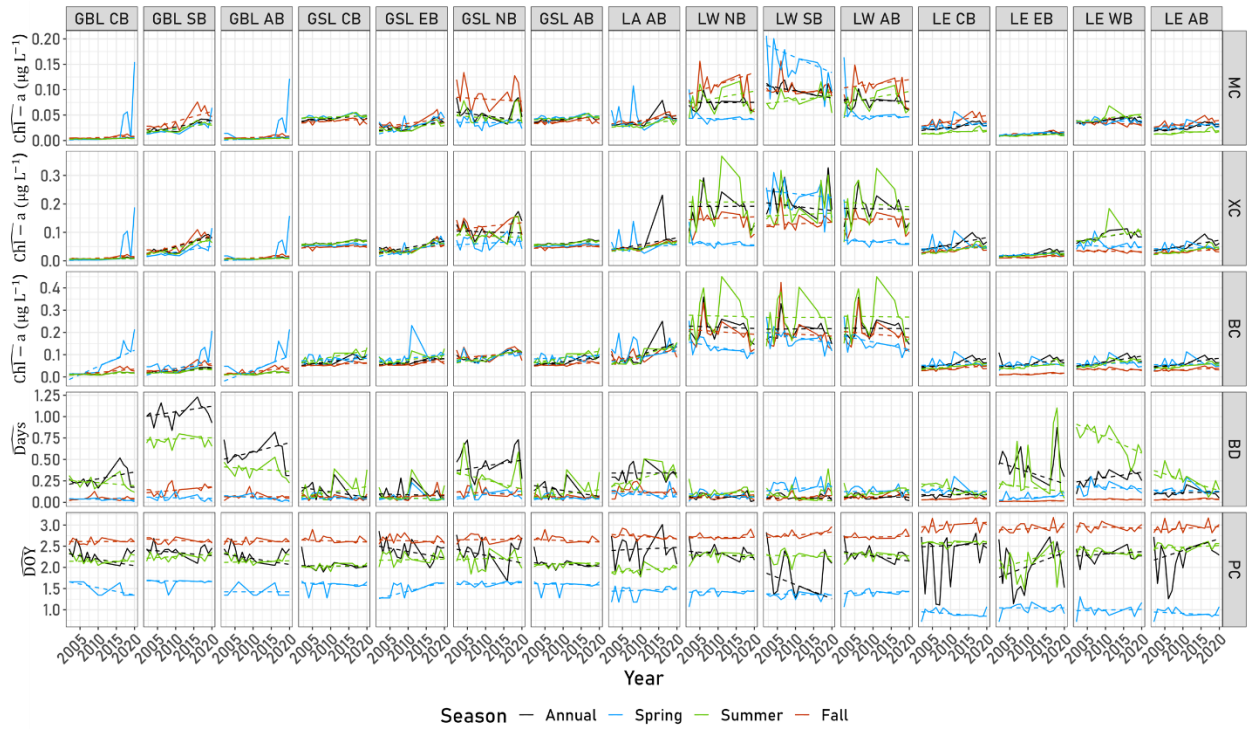
Figure 4.3. Time series of lake/basin-wide averages of 5 annual/seasonal algal biomass parameters for 15 lakes/basins. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All basins, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, and PC = Peak Chl-a DOY. All values are normalized to improve visualization across lakes/basins.
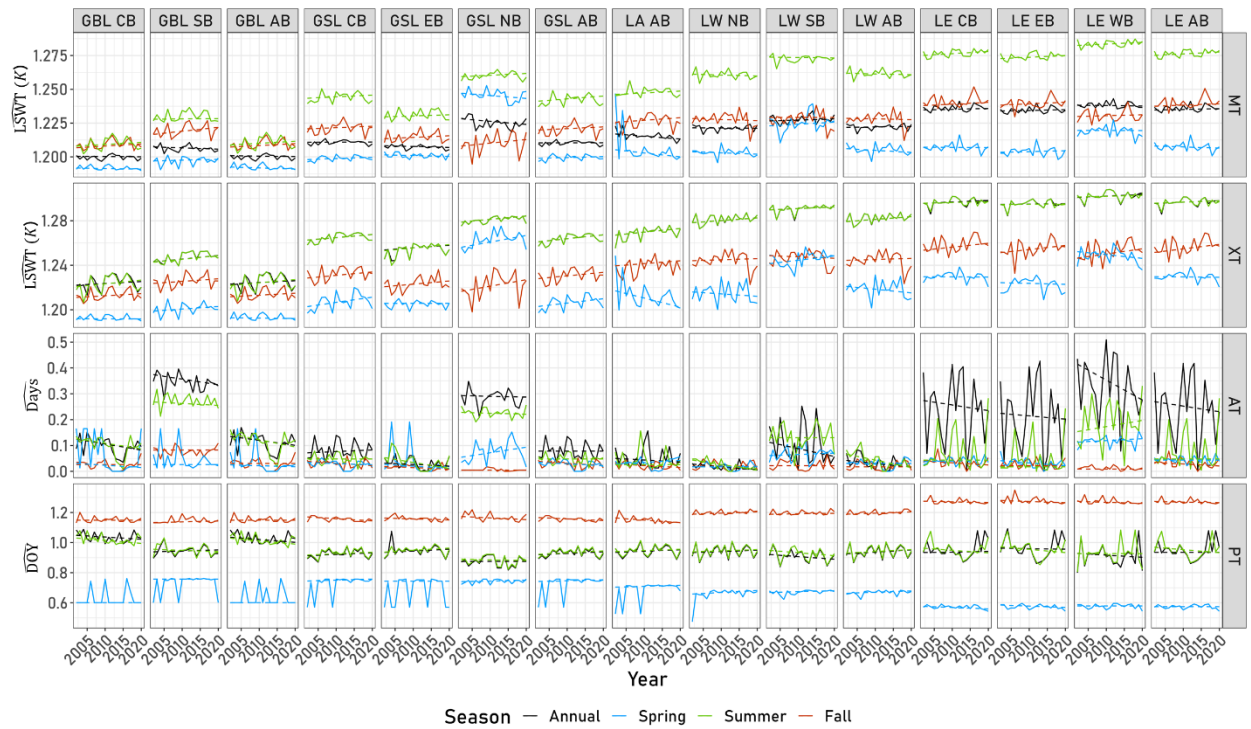
Figure 4.4. Time series of lake/basin-wide averages of 4 annual/seasonal LSWT parameters for 15 lakes/basins. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All basins, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin, WB = West Basin. MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, and PT = Peak LSWT DOY. All values are normalized to improve visualization across lakes/basins.
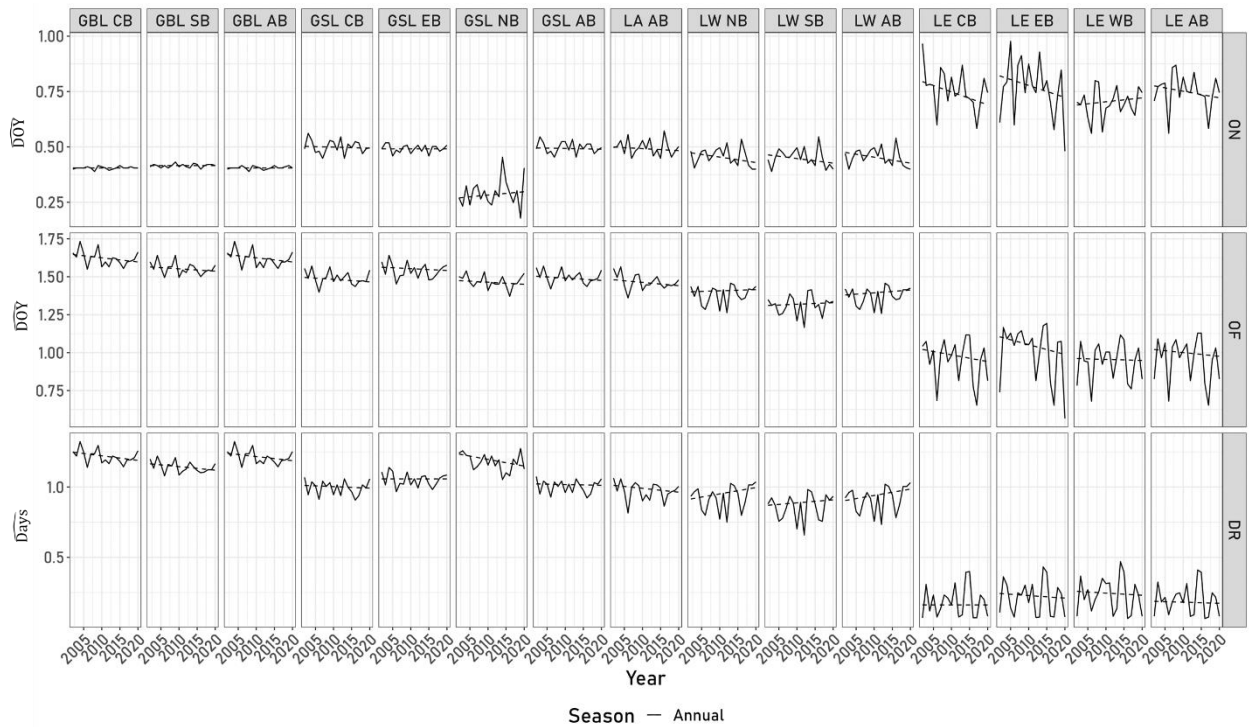
Figure 4.5. *T*ime series of lake/basin-wide averages of 3 annual/seasonal LIC parameters for 15 lakes/basins. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All basins, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin, WB = West Basin. ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration. All values are normalized to improve visualization across lakes/basins.

4.5.3 Variability of Timeseries

The variance of each algal biomass, LSWT and LIC parameter differed between seasons and lakes, as measured by the s_e of a Theil-Sen test. Of the MC, XC and BC parameters, SL typically exhibited higher variance ($\widetilde{s_e} = 1.96\,\mu g\,L^{-1}$) compared to the NL ($\widetilde{s_e} = 0.82\,\mu g\,L^{-1}$) across all seasons and parameters. The fall and spring seasons showed higher variances for all lakes ($\widetilde{s_e} = 2.02\,\mu g\,L^{-1}$) compared to annual and summer ($\widetilde{s_e} = 0.86\,\mu g\,L^{-1}$) seasons for MC, XC and BC, with the exception of LW which returned high variance for all seasons. LE frequently had the lowest variance for MC, XC and BC, with the exception of the fall season. Of the BD and PC parameters, there was little difference between variances observed for NL ($\widetilde{s_e} = 9.83\,days$) and SL ($\widetilde{s_e} = 7.78\,days$), with the exception of LE EB which observed a significantly higher variance compared to all other lakes in the summer season. Of the MT and XT parameters, NL observed higher variance in the summer

78

season ($\widetilde{s}_e = 1.00\ K$) compared to the SL ($\widetilde{s}_e = 0.71\ K$). For XT, spring and fall seasons observed higher variance ($\widetilde{s}_e = 1.41\ K$) compared to the annual and summer seasons ($\widetilde{s}_e = 0.99\ K$). Of the AT and PT parameters, NL observed higher variance in the spring season ($\widetilde{s}_e = 13.76\ days$) compared to SL ($\widetilde{s}_e = 3.81\ days$), while annual and summer seasons observed higher variance for SL ($\widetilde{s}_e = 16.52\ days$) compared to NL ($\widetilde{s}_e = 7.51\ days$). The fall season for AT and PT returned the lowest variance with no significant differences between lakes. Of the LIC parameters, SL observed higher variance ($\widetilde{s}_e = 21.13\ days$) compared to the NL ($\widetilde{s}_e = 8.71\ days$), particularly for LE (Figure 4.6).
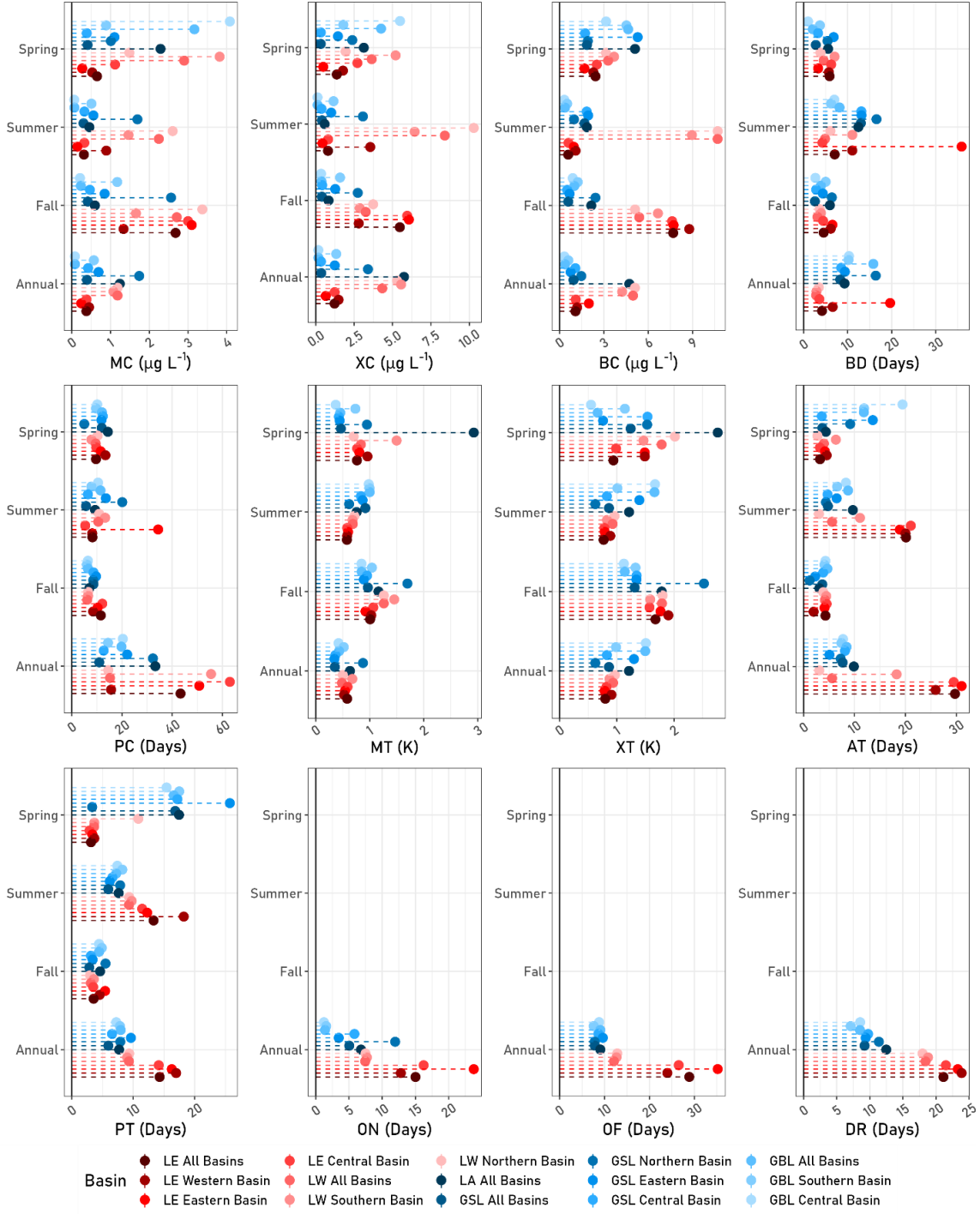
Figure 4.6. Theil-Sen standard error (s*e*) calculated for 15 lake/basins at an annual/seasonal scale for 12 parameters. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

4.5.4 Vector Autoregression Results

The VAR is used to determine how LSWT and LIC explained the trends and variability of the algal biomass response over time while accounting for the previous time-steps (lag). Lag selection was determined by the mean performance (adj. r^2, s_e and BIC) across all possible combinations of LSWT and LIC parameters for all seasons. The sample size ($n = 19$) is further reduced by the size of the lag ($n - lag$), which limits the number of endogenous variables to be included. For the MLR (lag = 0) and a VAR with a lag of 1, upwards of 8 input parameters were included in the VAR model (127 combinations), lag of 2 tested upwards of 5 input parameters (98 combinations) and a lag of 3 upwards of 3 input parameters (28 combinations). A lag order of 3 was selected as 2-4 of 5 total biomass parameters showed the highest mean adj. r^2 across the seasons, 4 of 5 showed the lowest s_e, and 5 of 5 showed the lowest BIC (Table 4.4).

Table 4.4. Mean performance and error of MLR (0 lag) and VAR models (1-3 lag order) for each algal biomass parameter. MC represents mean chl-*a*, XC is maximum chl-*a*, BC is abnormally high chl-*a*, BD is abnormally high chl-*a* days, and PC is maximum chl-*a* DOY. Bold indicates the best-performing metric for each algal biomass parameter and season.

Parameter		Adj. r2					s_e					BIC				
		MC	XC	BC	BD	PC	MC	XC	BC	BD	PC	MC	XC	BC	BD	PC
Annual	MLR	-0.01	0.00	-0.04	0.07	0.09	1.00	1.00	1.02	0.96	0.94	65.02	64.84	65.42	63.02	62.38
	1	0.43	0.49	**0.62**	**0.14**	**0.04**	0.73	0.68	0.58	**0.93**	0.85	52.03	48.79	42.60	61.66	57.72
	2	0.43	0.44	0.61	0.00	-0.06	0.72	0.71	0.58	1.01	0.88	50.34	49.65	41.72	63.32	57.84
	3	**0.49**	**0.49**	0.60	-0.02	-0.01	**0.66**	**0.65**	0.58	1.01	**0.81**	**43.52**	**42.71**	**38.71**	59.25	51.65
Spring	MLR	0.04	0.05	0.07	0.09	0.09	0.97	0.96	0.96	0.95	0.94	63.08	62.74	62.99	62.61	61.89
	1	0.03	0.08	0.13	0.12	0.17	0.87	0.83	0.86	0.92	0.79	56.28	55.42	58.28	60.84	55.38
	2	0.06	0.01	0.12	0.18	0.21	0.87	0.85	0.86	0.88	**0.76**	54.66	55.50	56.49	57.90	53.24
	3	**0.16**	**0.13**	**0.19**	**0.24**	0.09	**0.76**	0.77	0.78	0.84	0.78	**45.45**	47.87	48.52	52.37	50.40
Summer	MLR	0.02	0.02	0.06	0.03	0.04	0.99	0.99	0.97	0.98	0.98	64.41	64.41	63.64	63.95	63.74
	1	0.37	0.52	0.56	**0.21**	0.00	0.78	0.66	0.63	**0.85**	0.96	54.76	48.06	46.64	57.85	63.00
	2	0.37	0.52	0.58	0.18	-0.02	0.79	0.67	0.62	0.88	0.94	54.56	48.37	45.59	58.22	60.55
	3	**0.46**	**0.56**	**0.61**	0.20	**0.13**	**0.73**	**0.63**	**0.57**	0.85	**0.87**	**48.62**	**43.08**	**39.12**	53.33	53.89
Fall	MLR	0.03	0.01	0.01	-0.02	0.08	0.98	0.99	0.99	1.01	0.95	64.17	64.58	64.55	65.11	62.86
	1	**0.24**	0.33	0.34	0.06	**0.09**	0.68	0.61	0.57	0.93	0.94	44.31	39.70	36.69	61.42	62.28
	2	0.19	0.31	0.31	0.04	0.13	0.71	0.61	0.57	0.94	**0.92**	45.54	40.16	36.06	60.74	59.87
	3	0.23	**0.35**	**0.35**	**0.07**	0.03	**0.67**	**0.58**	**0.53**	**0.89**	0.98	**41.21**	**36.28**	**31.50**	55.25	57.99

To determine the importance of LIC in driving the trends and variability in annual/seasonal algal biomass, the mean ranking of performance/error metrics for VAR models (Table 4.2) containing only LIC, LSWT, and both LIC and LSWT parameters were assessed. LSWT only VAR models consistently returned higher performance for NL ($\overline{rank} = $ 13.31) for all seasons, with the exception of spring, in which LIC only models were higher performing ($\overline{rank} = $ 13.31). LIC only models in NLs returned the highest ranking in the majority of lake basins for the MC parameter, during the annual, spring and fall seasons (GSL CB, GSL NB, GSL AB and LA AB) (Table 4.5). For XC and BC models, LSWT only models consistently returned a higher performance ($\overline{rank} = $ 13.85) compared to LIC parameters ($\overline{rank} = $16.06). The performance of LSWT only, LIC only and a mix of both varied between lakes for the BD and PC parameter, showing no discernible pattern.

Table 4.5. Number of lakes/basins with the highest ranking for VAR models consisting of only LIC, LSWT or both LIC and LSWT within Northern and Southern regions.

Season	Parameter	Northern Lakes			Southern Lakes		
		LIC	LSWT	Both	LIC	LSWT	Both
Annual	Mean Chl-*a*	4	4	0	1	6	0
	Max Chl-*a*	2	6	0	3	4	0
	Abnormal Chl-*a*	1	6	1	1	6	0
	Abnormal Chl-*a* Days	1	6	1	2	4	1
	Peak Chl-*a* DOY	1	5	2	4	2	1
Spring	Mean Chl-*a*	4	3	1	1	5	1
	Max Chl-*a*	2	5	1	2	5	0
	Abnormal Chl-*a*	1	7	0	3	4	0
	Abnormal Chl-*a* Days	2	3	3	3	3	1
	Peak Chl-*a* DOY	4	4	0	3	4	0
Summer	Mean Chl-*a*	3	5	0	2	5	0
	Max Chl-*a*	3	4	1	3	4	0
	Abnormal Chl-*a*	1	6	1	4	2	1
	Abnormal Chl-*a* Days	3	3	2	5	2	0
	Peak Chl-*a* DOY	2	4	2	2	5	0
Fall	Mean Chl-*a*	4	3	1	3	4	0
	Max Chl-*a*	4	3	1	3	4	0
	Abnormal Chl-*a*	1	5	2	1	6	0
	Abnormal Chl-*a* Days	1	4	3	3	4	0

Season	Parameter	Northern Lakes			Southern Lakes		
		LIC	LSWT	Both	LIC	LSWT	Both
	Peak Chl-*a* DOY	3	4	1	1	6	0

VAR Model performance for NL varied significantly across lakes/basins and seasons (Figures 4.7 and 4.8). NL annual VAR models for MC, XC and BC parameters provided adequate performance for MC, XC and BC models ($\widetilde{adj.\ r^2} = 0.75, s_e = 0.51$), while BD and PC returned poor results for all seasons ($\widetilde{adj.\ r^2} = 0.10, s_e = 0.90$). Spring VAR models for MC, XC and BC parameters returned the lowest performance of any season ($\widetilde{adj.\ r^2} = 0.40, s_e = 0.79$).

VAR Model performance for SL returned frequent poor results for LW, and adequate results for LE; however, frequently poorer in comparison to NLs. SL annual and summer VAR models for MC, XC and BC parameters provided adequate results ($\widetilde{adj.\ r^2} = 0.46, s_e = 0.70$), while BD and PC also returned poor results for all seasons ($\widetilde{adj.\ r^2} = 0.06, s_e = 0.84$). Most SL spring and fall VAR models for all algal biomass parameters returned poor p`erformances ($\widetilde{adj.\ r^2} = -0.02, s_e = 0.79$). Despite the poor performance on average, LW returned adequate spring VAR model performances for BC ($\widetilde{adj.\ r^2} = 0.52, s_e = 0.42$). For full results, refer to supplementary tables B.7.5 to B.7.8.
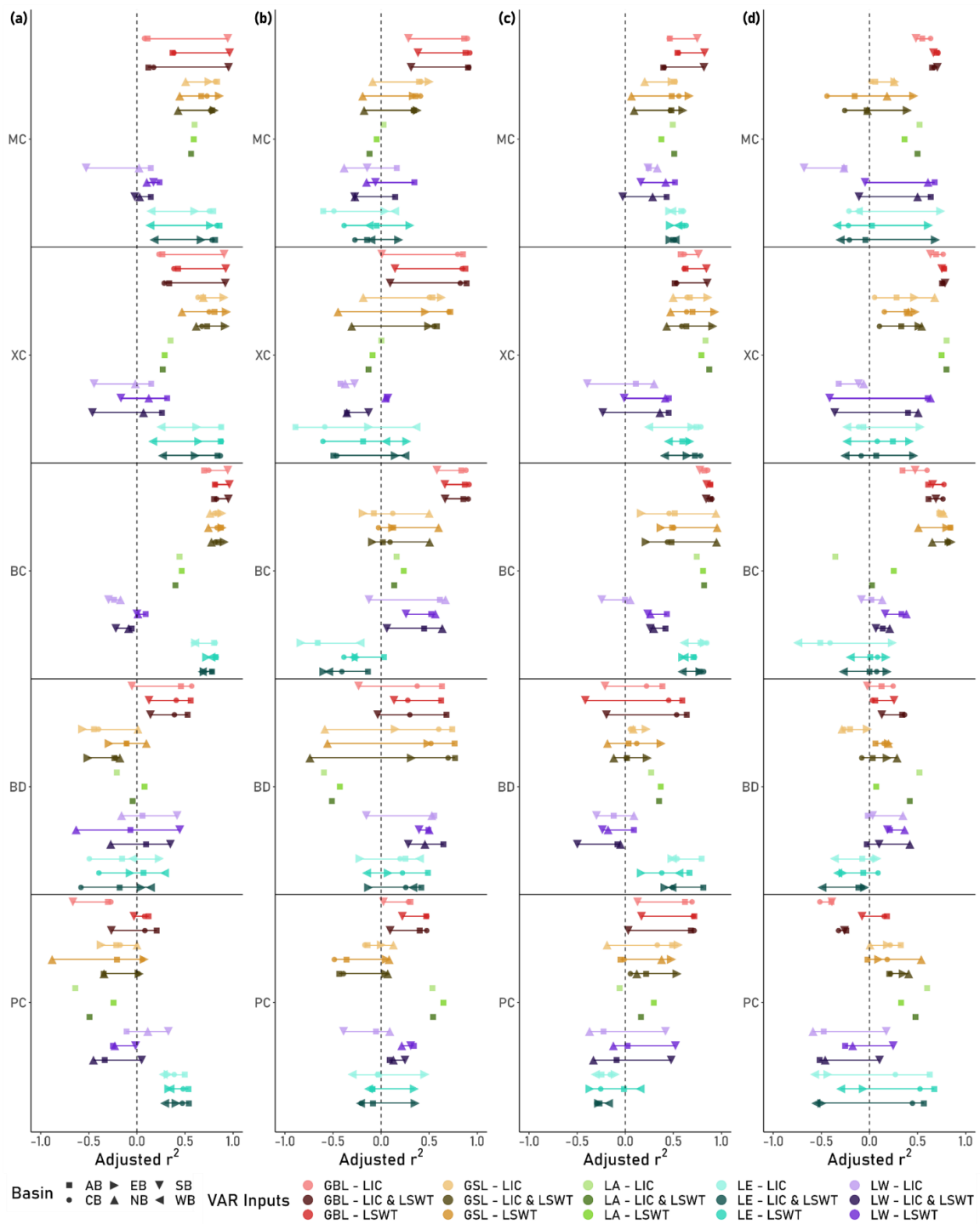
Figure 4.7. Plots representing VAR model mean adj. r^2 for five algal biomass parameters. Colour indicates the lake, shapes the basin and the shade whether the VAR models contained only LIC parameters, only LSWT parameters, or both for (a) Annual, (b) Spring, (c) Summer and (d) Fall.

84
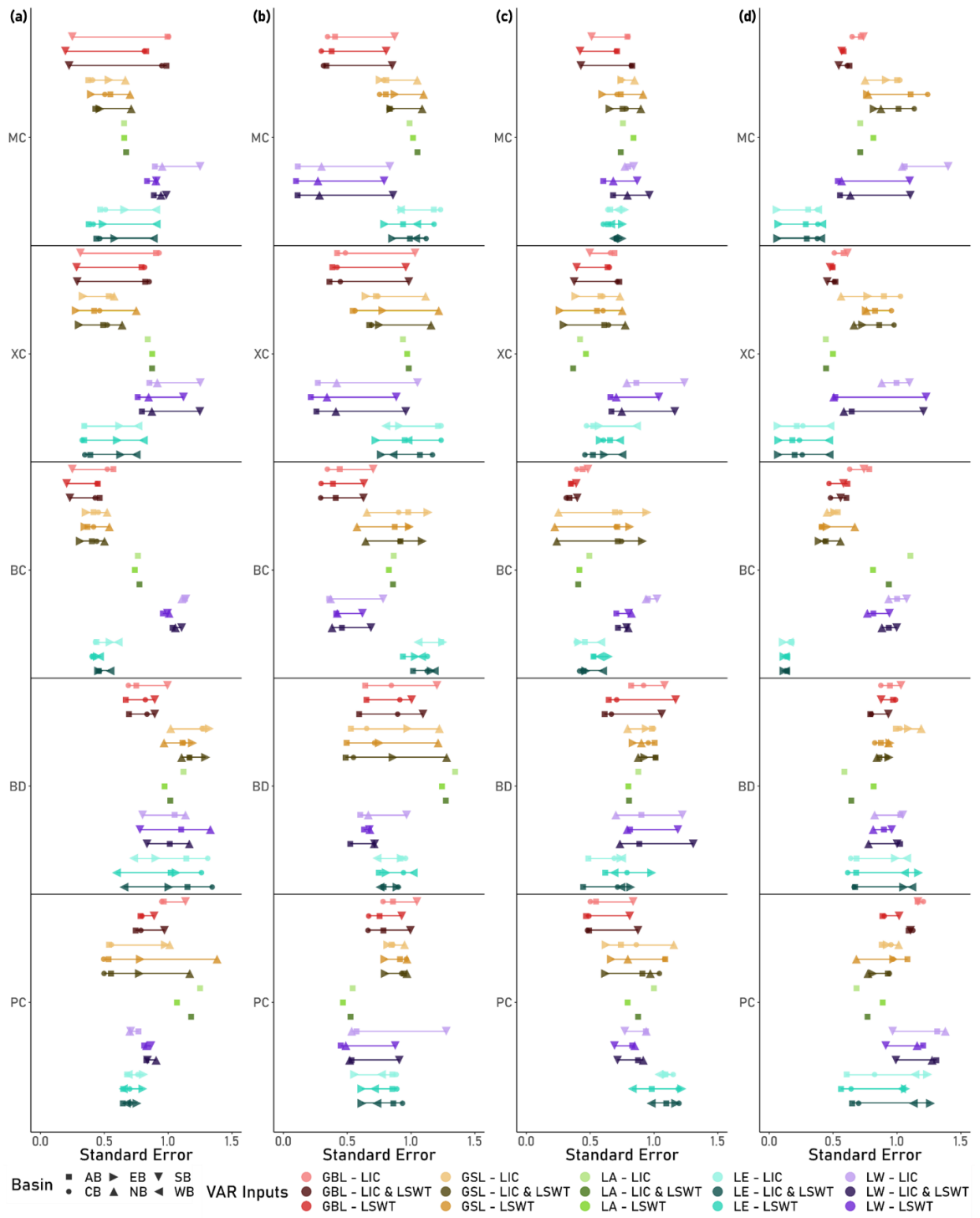
Figure 4.8. Plots representing the VAR model mean standard error (s_e) for five algal biomass parameters. Colour indicates the lake, shapes the basin and the shade whether the VAR models contained only LIC parameters, only LSWT parameters, or both for (a) Annual, (b) Spring, (c) Summer and (d) Fall.

85

4.5.5 Parameter Performance

VAR models were ranked based on the average rank of each performance metric (Table 4.2) when a given parameter was included. LSWT parameters were associated with the most frequent high performing VAR models across all biomass parameters and seasons for NL (Figure 4.9). The inclusion of MT and AT in annual NL VAR models returned a higher performance compared to that of other parameters for all algal biomass parameters ($\overline{rank} = 1.9$). While not found as the highest performing parameter, OF returned a higher performance when included in spring and summer NL VAR models across all algal biomass parameters ($\overline{rank} = 2.7$). The ON was frequently associated with the poorest performing NL VAR models across all algal biomass parameters and seasons ($\overline{rank} = 6.2$). For full results, please refer to supplementary tables B.7.5 to B.7.8.

Similar to the NL, LSWT were associated with the most frequent high performing VAR models across all biomass parameters and seasons for SL (figure 4.9). The inclusion of XT, AT and PT in annual SL VAR models returned a higher performance compared to that of other parameters for all algal biomass parameters ($\overline{rank} = 2.7$). Similar to the NL, LIC parameter performance increased for spring and summer SL VAR models, however, was less consistent across algal biomass parameters. The inclusion of ON for spring SL VAR models returned higher performance for BC, BD and PC ($\overline{rank} = 3.3$), while OF returned higher performance for XC, BC and BD in summer ($\overline{rank} = 2.0$). The inclusion of MT and XT in fall SL VAR models returned higher performance across all algal biomass parameters ($\overline{rank} = 2.2$). For full results, please refer to supplementary tables B.7.5 to B.7.8.
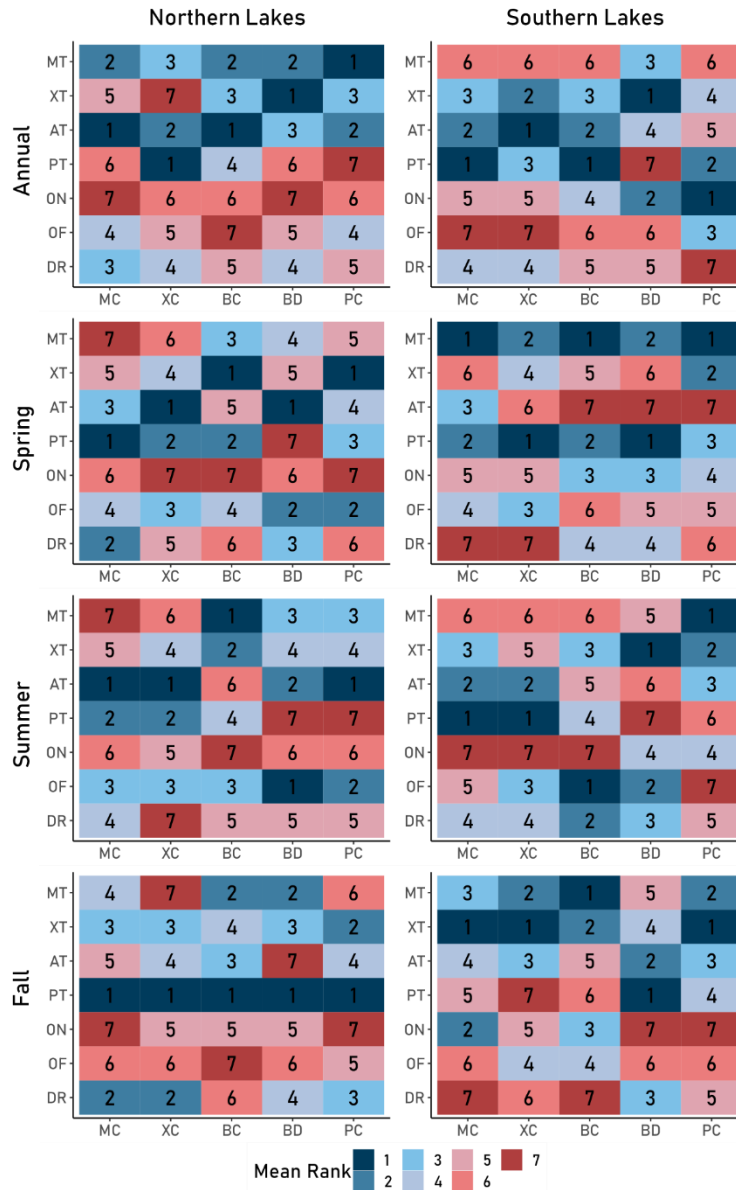
Figure 4.9. Ranking (1 to 7) of VAR model performance when a given independent variable was included. Model performance was determined by the calculating the median ordered rank of each VAR model performance metric: Corrected (adjusted) coefficient of determination (adj. r^2), Akaike information criterion second order (AICc), Bayesian Information Criterion (BIC), Standard Error (s_e) and final prediction error (fpe). The rankings were calculated separately for Northern Lakes and Southern Lakes using annual/seasonal data. The y-axis represents the independent variable, the x-axis the dependent variable. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

87

4.5.6 Model Performances

The VAR models with the highest ranking (Table 4.6) for NL explained -21% – 99% of the variance in all algal biomass parameters across all seasons ($\widetilde{adj.}\ r^2 = 0.78, s_e = 0.47$). Most of the top-ranked annual NL VAR models contained MT and AT parameters (45% and 40%, respectively) (Table 4.7), where the coefficients showed frequent positive β (supplementary figure B.7.2). However, the majority of the top-ranked spring and summer NL VAR models contained OF and XT parameters (40% – 43% and 35% – 40%, respectively), where VAR model coefficients returned a primarily negative β for OF in spring and positive in summer. The majority of top performing fall NL VAR models contained PT parameters (38%), which exhibited a primarily negative β with XC and BC, and positive with MC and BD parameters (supplementary figure B.7.2). Of the top-ranked NL VAR models, MC, XC and BC parameters returned a higher performance ($\widetilde{adj.}\ r^2 = 0.87, s_e = 0.38$) across all seasons. Conversely, the top-ranked NL VAR models returned a poorer performance for BD and PC parameters ($\widetilde{adj.}\ r^2 = 0.66, s_e = 0.53$) across all seasons.

The top-ranked VAR models (Table 4.6) for SL returned a poorer performance compared to the NL, which explained 10% – 99% of the variance in algal biomass parameters across all seasons ($\widetilde{adj.}\ r^2 = 0.68, s_e = 0.49$). The majority of top-ranked annual and fall SL VAR models contained primarily XT, AT and PT parameters for most lakes (31% – 34%, 31% – 34%, and 29% – 37%, respectively). Annual VAR model β coefficients were primarily negative when XT and PT were included, annual AT and fall XT, PT and AT were mixed depending on the lake/basin (supplementary Figure B.7.2). Similar to NL, LIC frequency had increased amongst the top-ranked spring and summer SL VAR models; however, MT and PT parameters remained the most frequent occurring (40 – 45.71% and 42.86%, respectively), where spring VAR model β coefficients were primarily positive when MT or PT were included; summer PT was primarily negative (supplementary Figure B.7.2). Of the top-ranked NL VAR models, MC, XC and BC parameters typically returned a higher performance ($\widetilde{adj.}\ r^2 = 0.71, s_e = 0.40$) across all seasons. The spring VAR MC and BC models had

however returned a poorer performance ($\widetilde{adj.\ r}^2 = 0.46, s_e = 0.68$) in comparison to the XC, BD and PC models ($\widetilde{adj.\ r}^2 = 0.71, s_e = 0.46$).

Table 4.6. Top-ranked annual VAR models for each lake and basin performance and error metrics. MT = mean LSWT, XT = Max LSWT, AT = anomaly LSWT days, PT = peak LSWT DOY, MC = mean chl-*a*, XC = max chl-*a*, BC = abnormal chl-*a*, BD = abnormal chl-*a* days, PC = peak chl-*a* DOY. Filled box indicates a p-value < 0.1, empty box indicates a p-value > 0.1.

	Model	Adj. r2	AICc	fpe	s_e	P		Model	Adj. r2	AICc	fpe	s_e	P
GBL Central Basin	MC ~ XT + AT	0.63	140.56	0.01	0.64	■	**LW North Basin**	MC ~ XT + PT	0.68	135.40	0.36	0.55	■
	XC ~ AT + DR	0.92	117.25	0.12	0.31	■		XC ~ XT + DR	0.54	139.35	0.53	0.62	□
	BC ~ MT + AT	0.92	115.36	0.01	0.29	■		BC ~ XT + PT	0.60	141.09	0.41	0.65	□
	BD ~ OF + DR	0.87	124.05	0.00	0.38	■		BD ~ MT + ON	0.69	137.80	0.81	0.59	■
	PC ~ AT + ON	0.74	127.84	0.01	0.43	■		PC ~ ON + DR	0.51	134.45	0.72	0.53	□
GBL South Basin	MC ~ AT	0.99	14.35	0.01	0.12	■	**LW South Basin**	MC ~ MT + ON	0.80	129.92	0.88	0.46	■
	XC ~ PT + ON	0.97	100.22	0.00	0.18	■		XC ~ MT + ON	0.46	146.27	0.67	0.77	□
	BC ~ AT	0.98	18.68	0.02	0.14	■		BC ~ XT	0.19	79.03	0.28	0.90	□
	BD ~ MT + AT	0.59	139.83	0.09	0.63	□		BD ~ XT + ON	0.79	131.63	0.25	0.48	■
	PC ~ XT + ON	0.58	136.91	1.45	0.57	□		PC ~ ON + DR	0.50	138.94	0.14	0.61	□
GBL All Basins	MC ~ MT + AT	0.81	129.87	0.08	0.46	■	**LW All Basins**	MC ~ PT + AT	0.58	139.91	0.03	0.63	□
	XC ~ AT + DR	0.92	117.03	0.05	0.31	■		XC ~ XT + PT	0.62	136.90	0.39	0.57	■
	BC ~ MT + AT	0.93	114.42	0.01	0.28	■		BC ~ XT + PT	0.64	139.11	0.36	0.61	■
	BD ~ MT + XT	0.84	125.57	0.00	0.40	■		BD ~ MT + AT	0.69	138.72	0.17	0.61	■
	PC ~ AT + DR	0.66	132.15	0.03	0.49	■		PC ~ AT + ON	0.77	121.20	0.27	0.35	■
GSL Central Basin	MC ~ PT + DR	0.97	100.32	0.09	0.18	■	**LE Central Basin**	MC ~ MT + AT	0.92	114.45	0.09	0.28	■
	XC ~ MT + DR	0.90	116.03	0.04	0.30	■		XC ~ MT + AT	0.99	82.26	0.01	0.10	■
	BC ~ MT + OF	0.91	116.80	0.02	0.31	■		BC ~ PT + AT	0.97	98.19	0.02	0.17	■
	BD ~ MT + OF	0.50	146.18	1.04	0.76	□		BD ~ XT	0.14	82.30	1.55	1.00	□
	PC ~ AT + OF	0.60	118.77	0.07	0.32	□		PC ~ MT + ON	0.68	136.06	0.35	0.56	■
GSL East Basin	MC ~ PT + ON	0.94	112.83	0.01	0.27	■	**LE East Basin**	MC ~ PT + AT	0.92	115.43	0.01	0.29	■
	XC ~ MT + PT	0.97	99.91	0.01	0.18	■		XC ~ PT	0.73	61.60	0.06	0.52	■
	BC ~ MT + AT	0.95	108.12	0.01	0.23	■		BC ~ AT + DR	0.95	101.93	0.02	0.19	■
	BD ~ XT + PT	0.22	152.66	0.07	0.94	□		BD ~ ON + OF	0.44	146.63	0.02	0.77	□
	PC ~ AT + OF	0.81	122.78	0.43	0.37	■		PC ~ XT + ON	0.77	131.01	0.01	0.48	■
GSL North Basin	MC ~ AT + DR	0.63	137.04	0.05	0.57	■	**LE West Basin**	MC ~ MT + DR	0.44	144.98	0.29	0.74	□
	XC ~ PT + DR	0.82	128.50	0.19	0.44	■		XC ~ PT + DR	0.65	134.18	0.08	0.53	■
	BC ~ XT + DR	0.84	127.59	0.02	0.43	■		BC ~ PT + AT	0.82	126.44	0.08	0.41	■
	BD ~ MT + AT	0.63	139.67	1.24	0.62	■		BD ~ MT + AT	0.60	129.47	0.16	0.45	□

Group	Model						Group	Model				
	PC ~ ON	0.11	80.89	2.55	0.96	□		PC ~ XT + OF	0.77	125.04	0.06	0.39	■
GSL All Basins	MC ~ MT + OF	0.96	102.71	0.02	0.20	■	LE All Basins	MC ~ PT	0.91	44.31	0.02	0.30	■
	XC ~ MT	0.87	48.32	0.03	0.35	■		XC ~ MT + AT	0.97	99.52	0.13	0.18	■
	BC ~ AT + ON	0.90	118.57	0.05	0.32	■		BC ~ PT	0.88	47.90	0.05	0.34	■
	BD ~ MT + OF	0.53	144.90	1.30	0.73	□		BD ~ XT + DR	0.48	146.45	0.24	0.77	□
	PC ~ AT + OF	0.59	117.79	0.03	0.31	□		PC ~ PT	0.71	61.45	0.59	0.52	■
LA	MC ~ MT	0.76	60.77	0.06	0.51	■							
	XC ~ MT	0.55	70.88	0.14	0.70	■							
	BC ~ MT	0.71	63.13	0.08	0.55	■							
	BD ~ PT + DR	0.75	133.37	0.42	0.51	■							
	PC ~ MT + XT	0.62	138.39	0.05	0.60	■							

Table 4.7. Frequency of predictor parameter (MT = mean LSWT, XT = Max LSWT, AT = anomaly LSWT days, PT = peak LSWT DOY) presence in top-ranked VAR models within Northern (NL) and Southern Lakes (SL).

	Annual		Spring		Summer		Fall	
Parameter	NL	SL	NL	SL	NL	SL	NL	SL
MT	45%	29%	13%	46%	25%	40%	23%	31%
XT	15%	31%	35%	14%	40%	31%	33%	34%
AT	45%	31%	28%	11%	33%	29%	28%	34%
PT	18%	37%	28%	43%	25%	43%	38%	29%
On	15%	29%	10%	17%	8%	6%	33%	26%
OF	20%	6%	40%	20%	43%	17%	18%	14%
DR	25%	20%	35%	37%	8%	31%	28%	17%

4.5.7 Model Variance

When looking at the VAR model residuals, it was observed that MC, XC and BC concentration models returned low variance in GBL SB and the LE EB. BC also returned low variances for all GBL, GSL and LE compared to the other lakes and algal biomass parameters. Of the MC, XC and BC parameters, models showed increased variance over time in GBL, LA and, LE particularly, around 2016-2020, while LW showed an inverse trend for BC. BD and PC showed greater variability in all lakes (Figure 4.10).

Figure 4.10. VAR model residuals for five algal biomass parameters, where the colour represents the lake basin.

4.6 Discussion

4.6.1 Pre-Processing and Sources of Error

Daily chl-*a* concentrations at a 1-km grid provided an opportunity to understand the phenology of algal biomass in many large freshwater lakes. Variance due to retrieval error introduced a challenge for data analysis, where pre-processing was implemented to mitigate retrieval error issues. Missing data due to cloud cover, sensor image gaps, and outliers (*i.e.*, retrieved chl-*a* concentrations of > 500.00 μg L^{-1} in oligotrophic lakes) required interpolation. While more complex interpolation methods exist, linear temporal interpolation can provide a fast, simple and reliable means of interpolating smaller missing data gaps (Oehmcke *et al.*,

91

2016; Schlegel *et al*., 2019). Missing data from 2012-2015 due to a gap in sensor image archives (2002-2011 for MERIS, 2016-present for Sentinel-3A), limited the capacity of the VAR models to identify trends due to interpolation. The data derived from Sentinel-3A (2016 – 2020) typically returned higher chl-*a* concentration for primarily GBL, LA and LE compared to MERIS (2002 – 2011), presenting potential sensor bias as seen in the time series (Figure 4.3) and the variance plots (Figure 4.10). This sensor bias may influence the trends observed, providing a stronger trend then is accurate. Future versions of the ESA CCI Lakes product may improve these findings. As the jump in variance was not observed for all lakes/basins, the trends may instead be a result of shifting algal biomass dynamics; however, higher temporal resolution is needed (Seekell *et al*., 2011). The ice on/off period may have introduced higher error, due to a reduction of available open water pixels, mixed pixilation of ice-water edges, or increased turbidity due to higher runoff, resulting in an over estimation of chl-*a* concentrations (Jiang *et al*., 2023; Zhang & Pavelsky, 2019). Overestimation of chl-a concentration in turbid waters is a common issue in optical remote sensing, in which the application of OWT was employed to minimize this issue (Carrera *et al*., 2023). ON date identification was also limited due to high cloud coverage during the ice-forming season, however, was mitigated using interpolation methods.

Seasonal thresholds were fixed across all years to maintain consistency in data separation across all years, however seasonality can be transient. Warm spring seasons will have earlier growing periods, while some years may be cooler and start later. A fixed threshold impacts the number of available pixels for each year, where warmer spring periods have a greater number of open water pixels, while colder periods have fewer. Higher ice coverage can impact chl-*a* retrieval due to the adjacency effect and mixed pixelation (Jiang *et al*., 2023).

4.6.2 Trend Analysis of Input Parameters

Despite the increase in chl-*a* concentrations from Sentinel-3A retrieved data, most lakes (GSL, LA and LE) showed a positive trend in MC, XC and BC from 2002-2011 (Figure 4.3). The increase in chl-*a* concentrations coincides with other studies which found both an increase and decrease in biomass (*i.e*., LW) (Ho *et al*., 2019). Kraemer *et al*. (2022) found that

63% of the 344 globally distributed lakes studied had an increasing trend in chl-*a* concentrations (1997-2020); however, only 44% of the lake surface area showed a positive trend. Lakes/basins in this study are treated as a singular value, therefore it is unknown if this trend persists across the majority of the lake surface area. Studies have shown that eutrophic lakes typically exhibit a positive trend in increasing algal biomass with increasing LSWT, compared to oligotrophic lakes which decreased (Kraemer *et al*., 2017). Oligotrophic lakes have also shown a decrease in pH and subsequent algal biomass due to atmospheric carbon deposition due to the formation of carbonic acid, while lakes with higher algal biomass increase pH and showed higher growth rates (Talling, 1976; Verschoor *et al*., 2013; Raven *et al*., 2020; Li *et al*., 2016). It stands to reason that in large lakes with complex morphometries (such as the study lakes), regions with higher nutrient inflow via non-algal particles would exhibit a positive trend but represent a smaller lake surface area. The rate of chl-*a* concentration increase likely outweighs the decrease, resulting in annual/seasonal spatial mean increases observed in MC, XC and BC (Figure 4.3), with the exception of LW, which showed a negative trend in algal biomass (supplementary Tables B.7.1 to B.7.4).

BC represents the mean of chl-*a,* which exceeded the defined threshold (see methods section 4.4). This represents chl-*a* concentrations greater than expected in a normal distribution for that entire lake surface area over a 19-year duration. The BD is a measure of the mean frequency that BC is present. Despite the increase in abnormal chl-*a*, the frequency decreased over time (GSL, LW and Eastern basin of LE), potentially indicating these abnormal surges in biomass are more intense but shorter lasting. The timing of the chl-*a* peak occurred earlier in the year except for LE, which may result from the earlier OF seen in most lakes. Earlier OF periods result in earlier internal loading and light availability, initiating earlier algal growth (Jewson *et al*., 2009). Lakes with limited external loading during the growing season (typical in remote regions) may see nutrient limitation later in the year due to the earlier open water season (Sommer *et al*., 2012), resulting in earlier algal biomass peaks. LE experiencing later peak chl-*a* is likely due to a longer open water season with higher nutrient replenishment and more frequent mixing events (Kane *et al*., 2014).

The mean observed annual XT returned an increasing trend, a result found in other studies (O'Reilly *et al*., 2015). However, annual/spring MT had a slight decreasing trend, but a slight increase in the summer and a more substantial increase in the fall (Figure 4.4). Due to the use of a fixed window during the spring season, later OF periods return fewer open water samples. Later OF periods often exhibit a higher rate of temperature change (Feiner *et al*., 2022; Dugan, 2021; Favot *et al*., 2019), where a smaller number of observations and a faster rate of LSWT change would likely return a higher mean, despite lower LSWT for the spring season. Due to the nature of remote sensing data, under ice water temperature is not resolved, and therefore not included in the mean calculations. Under ice LSWT is an assumed static 273.15K at the ice-water interface, however this temperature will vary due to snow depth, ice thickness, and ice composition (Palshin *et al*., 2019; Tsvetova, 2016), likely fluctuating between 273.15K and 277.15K (Huo et al., 2023).

4.6.3 VAR Models

Assessment of individual parameter trends provided insight into their variability and how they change over time; however, we would also like to know how the LIC and LSWT parameters interact and affect the algal biomass over time, particularly what the role of LIC is in algal biomass dynamics, as it is not fully understood. The LIC parameters were expected to explain a greater portion of the variance in algal biomass, particularly MC and XC in the NLs, compared to the lakes with higher anthropogenic influence. Lake ice facilitates the timing of internal loading and mixing, a period where oligotrophic lakes such as GBL and GSL have the highest available suspended nutrient concentrations for algal biomass to fixate (Sommer *et al*., 2012). While LIC parameters ranked higher in the NLs than in SLs, it was highly variable. LSWT plays a significant role in the growth rate of algae (Gobler *et al*., 2020) and typically explained a higher proportion of the variance, particularly during the summer. Despite ranking higher for MC and XC, the difference in performance compared to LSWT is negligible. LIC parameters were expected to be significant in explaining the variance of algal biomass in the spring season compared to others. However, LIC was more prevalent in spring for the NLs than the SLs; NLs did not have a higher variance explanation than other seasons. The poor

performance of VAR models for SLs during spring may be due to frequent overwintering snow/ice melt or fertilizer timing, which may explain the variance to a greater degree (Michalak *et al*., 2013).

Model performance typically worked well; however, GSL NB, LW SB and LE WB typically had a poorer performance, as LIC and LSWT did not explain a significant portion of the variance in algal biomass. Both the SB of LW and the west basin of LE are highly eutrophic (Binding *et al*., 2018; Binding *et al*., 2019) and near major river inflows within high agricultural regions. Frequent flushing, mixing and nutrient run-off due to precipitation and periodic crop fertilization are likely to trigger changes in algal biomass. Therefore, physical controls on internal loading are less critical. The NB of GSL is shallower compared to the rest of the lake, with a longer DR and higher water temperatures compared to the CB (Figures 4.7 and 4.8). The NB also has a significant legacy of mining-related contamination, of which has been observed to affect algal assemblages (Chételat *et al*., 2019), where it has been observed that sub-arctic lakes near Yellowknife had experienced trophic shifts due to eutrophication driven by urbanization and mining activities (Sivarajah *et al*., 2021), with Yellowknife Bay experiencing an AB in 2013 (Pick *et al*., 2016). It is possible that the north basin of GSL could be experiencing higher than baseline external loading compared to the other NLs. Algal community compositions have indicated that a regime shift had occurred post 2000CE within the CB of GSL, attributed to ice loss and rise in air temperatures (Rühland *et al*., 2023). The shallow NB of GSL may have experienced a stronger shift in biomass; however, further research is required to unravel this. Retrieval pre-processing errors are also potential sources for variance in these lakes as well.

Performances of individual parameters varied depending on season and lake/basin and followed what was expected with some exceptions. ON showed little change throughout most lakes and had difficulties in identification due to high cloud cover, which may also impact DR calculations. However, ON identification was mitigated by interpolation methods, providing estimations of ice formation occurrence between dates of observed open water and ice cover pixels for each year. ON was also associated with higher overall fpe and s_e, and was present in

fewer high-performing models, which is expected. OF and DR were most frequently found in the best performing spring VAR models, primarily for NLs, while SLs returned high performance for DR in only the WB of LE (which is unexpected due to the high variance of LIC for LE).

The summer season VAR model coefficients indicated that when OF or DR were included in a model, there was a frequent positive slope with MC, XC and BC parameters, indicating that a later OF or longer DR returned a decrease in chl-*a* concentrations in the summer (supplementary Figure B.7.2). It has been observed in other north temperate lakes that earlier OF and lower DR returned lower summer algal biomass (Hrycik *et al*., 2021). As a dimictic but shallow lake, the timing between OF and stratification due to high temperatures can lead to significant late summer/early fall ABs, as seen in Dickson Lake (located in Algonquin Provincial Park, Ontario), where a late ice melt coupled with rapid stratification resulted in incomplete mixing and a subsequent AB (Favot *et al*., 2019). Shallow, turbid waters stratify faster and may lead to poor mixing (Butcher *et al*., 2015; Coats *et al*., 2006). While the NLs are deep, some basins such as the SB of GBL and the NB of GSL remain quite shallow, where rapid stratification onset may be impactful. The duration of PT was the best predictive parameter (MC, XC and BC) on average for NLs and was consistently the highest ranked (Figure 4.9). This is possibly a result of the legacy effect high-temperature events have on deep lakes (Woolway & Merchant, 2018), where the lag order considers the impact of anomaly LSWT duration from prior years. While these parameters worked well for explaining the variance in changes to the concentration of chl-*a*, there was a generally poor performance in explaining the variance in the timing of peak chl-*a* and the duration of BC. There is considerable variability in the timing of chl-*a* peaks, as most lakes will exhibit multiple annual peaks, introducing variance in the VAR models. BD may also depend on the nutrient availability, quantity and frequency of external loading, as studies have found nutrient availability positively correlated with AB duration (Binding *et al*., 2018). Despite issues with some algal biomass parameter retrieval, VAR model performance with both LIC and LSWT

parameters provided a strong explanation of variance for annual time series of algal biomass parameters across all lakes. The NL VAR models however, did outperform those of the SLs.

4.7 Conclusion

LIC is an often-overlooked factor regarding the phenology of algal biomass despite its control on various lake properties, such as the timing of internal loading, mixing classification, and open water season. This study looked to determine if LIC could explain the trends and variability seen in algal biomass parameters (MC, XC, BC, BD and PC) in five North American Great Lakes. Algal biomass concentrations were found to increase annually, while LIC was decreasing and maximum LSWT was increasing. A VAR was used to establish a multivariate relationship between LIC/LSWT parameters and the algal biomass response to determine how these parameters interact over time. It was observed that NLs (GBL, GSL and LA) saw LIC parameters explain a greater portion of the variance in algal biomass trends (2002-2020) compared to SLs (LW and LE). NLs also had a greater explanation of variance in algal biomass trends during spring on average compared to SLs, which returned poor results. On average, NLs returned stronger VAR models than SLs, regardless of input parameters.

This research showcases some of the first uses of multiple ESA CCI Lakes variables in monitoring and assessing changes in lake phenology and biota. The current version of the ESA CCI Lakes product presents some limitations due to missing data (2012 – 2015), turbid waters during the OF period effecting chl-*a* retrieval, water-pixel identification and classification, amongst others, where future versions look to improve retrieval, uncertainty levels, and filling missing data gaps. Version 2.0.2 of the ESA CCI Lakes+ product does provide filled data gaps with MODIS aqua for some lakes, however they were excluded from this study due to abnormally high chl-a concentrations found during the MODIS years, likely due to high sensor bias. As future versions further improve in accuracy and scope, the methods/framework presented in this paper can be revisited to further improve on the results presented. Such methods can be applied to other lakes, as the ESA CCI Lakes+ product provides data for over 2000 large lakes globally. These results provide evidence that climatic parameters show significant explanatory power in NL systems that have limited anthropogenic influence. Under

a changing climate with exacerbated trends at higher latitudes, we can expect changes in algal biomass concentrations, timings and community compositions. While these results show explanatory power, predictive power is not as well understood. Additional research into how LIC, alone or in combination with LSWT, can predict algal biomass requires increased training observation data and potentially more robust prediction methods such as machine/deep learning techniques. Future research also aims to better understand the interaction effects of additional lake physical and atmospheric forcings on algal biomass dynamics in the same lakes. Such research can improve algal forecasting models, particularly for remote NLs.

Chapter 5: Is Lake Ice a Predictor of Algal biomass in North American Great Lakes?

5.1 Introduction

There has been a perceived increase in algal bloom frequency and severity globally (Favot *et al.,* 2023; Winter *et al.,* 2011; Paerl & Huisman, 2008), with occurrences increasing in recent years compared to past decades (Hou *et al.,* 2022). Climate change has often been indicated as a potential driver of this trend, since traditionally oligotrophic lakes (Favot *et al.,* 2019; Pick, 2016; Winter *et al.,* 2011; Smol, 2019) and high-latitude lakes (Ayala-Borda *et al.,* 2021; Freeman *et al.,* 2020) have been observed to bloom despite little to no anthropogenic activities. The primary impacts of climate change on algal biomass are attributed to increasing water temperatures (Borowitzka *et al.,* 2016; Cross *et al.,* 2015), shorter and more intense precipitation events followed by drought conditions (Reichwaldt & Ghadouani, 2012), and prolonged lake stratification periods (Jirsa *et al.,* 2013; Mosley, 2015). However, changes in lake ice cover (LIC) are often an overlooked mechanism in algal bloom formation and accumulation of algal biomass for lakes forming ice. LIC is altered by warming climate, with high latitudes exhibiting the highest rate of air temperature increase, particularly during the winter months (Hansen *et al.,* 2006; Houze *et al.,* 2019). LIC plays an important role in lake ecology and biogeochemical cycling, including the internal loading rate (Reynolds, 1994; Joung *et al.,* 2017), lake mixing classification (Vincent *et al.,* 2012), light availability (Hébert *et al.,* 2021) and algal community composition and succession (Salmaso *et al.,* 2018; Blenckner *et al.,* 2007). Additionally, LIC controls the duration of the algal growing season, where prolonged stratification due to a longer growing season has been observed to contribute to an increase in the prevalence of algal blooms (Jirsa *et al.,* 2013; Mosley, 2015), particularly for various cyanobacterial genus (Brasil *et al.,* 2016; Lehman *et al.,* 2017; Rühland et al., 2023). As a result, spring and summer algal biomass has been seen to increase following earlier ice-off periods in various studies (Adrian *et al.,* 1999; Preston *et al.,* 2016; Beyene & Jain, 2020). Algal biomass has also been observed to have increased in some European lakes following shorter ice cover periods (Weyhenmeyer *et al.,* 2008; Noges *et al.,* 2010). It has previously

been observed how the variability and trends in algal biomass relate to those of the LIC and the lake surface water temperature (LSWT) parameters (chapter 4), where LIC provided a greater explanation of the variance in algal biomass within Northern remote lakes such as Great Bear Lake and Great Slave Lake compared to more anthropogenically developed regions. However, it remains unclear whether LIC parameters are a strong predictor of algal biomass.

The impacts of LIC on to algal biomass dynamics is less understood compared to climatic parameters, as few studies have done direct comparisons. Some recent publications have included LIC in algal bloom modelling (Ai *et al.,* 2023; Lin *et al.,* 2023), where machine learning algorithms that included the timing of lake ice (on/off) had improved performance in the prediction of abnormally early spring algal blooms (Lin *et al.,* 2023). Additional studies have found that spring algal blooms were highly sensitive to warming in lakes controlled by ice-off/onset of stratification (Gronchi *et al.,* 2021). Others have found that reduced ice cover led to greater zooplankton productivity, resulting in increased bloom suppression (Hébert *et al.,* 2021). These recent studies have used increasingly open and available reanalysis and remote sensing data products to improve our understanding of algal biomass dynamics in a changing climate.

This research uses remote sensing data products provided by the European Space Agency's (ESA) Climate Change Initiative (CCI) Lakesproject to identify the importance of LIC in predicting annual and seasonal (spring, summer and fall) algal biomass parameters in five North American Great Lakes; Great Bear Lake, Great Slave Lake, Lake Athabasca, Lake Winnipeg, and Lake Erie, from 2002 to 2020 (Carrea *et al.,* 2022). The Lake Water Leaving Reflectance (LWLR) product provides daily gridded chlorophyll-*a* (chl-*a*; proxy of algal biomass) concentrations, which was used to calculate five annual and seasonal algal biomass parameters; mean (MC) and max chl-*a* (XC) concentrations, abnormally high chl-*a* concentrations (BC), abnormally high chl-*a* days (BD), and peak chl-*a* DOY (PC). The ESA CCI Lakes product provides harmonized lake data, where the LIC parameter was used to calculate lake ice on (ON), ice off (OF) and lake ice duration (DR). To determine the importance in LIC for algal biomass parameter prediction, the Lake Surface Water

Temperature (LSWT) product from the ESA CCI Lakesproject was used to calculate four annual and seasonal parameters; mean (MT) and max LSWT (XT), anomaly LSWT days (AT), and peak LSWT DOY (PT). LSWT is an important factor in the algae growth cycle and can therefore be compared with that of LIC to indicate where and when LIC is an important predictive factor in annual and seasonal algal biomass parameters. This research looks to determine if the LIC and LSWT parameters can predict for annual and seasonal algal biomass utilizing linear (multiple linear regression) and non-linear (feed-forward artificial neural networks) functions. Additionally, a random forest (RF) classifier was used to determine if the LIC and LSWT parameters can classify the presence of annual and seasonal BC (only pixels with a chl-a concentration z-score of > 2 of each given year). Due to the control on internal loading and light availability, it is hypothesized that LIC will provide a strong predictive performance for remote Northern Lakes (NL), particularly during the spring season, due to limited external nutrient loading. Furthermore, it is anticipated that LSWT will provide greater importance during the summer and fall periods compared to LIC. An improved understanding of how LIC and LSWT parameters can be used to predict algal biomass may improve future algal biomass projection models.

5.2 Study Area

Annual and seasonal algal biomass, LIC and LSWT parameters were calculated for five North American Great Lakes; Great Bear Lake (GBL), Great Slave Lake (GSL), Lake Athabasca (LA), Lake Winnipeg (LW) and Lake Erie (LE) and their subsequent basins from 2002-2020. This research utilizes the same study area as outlined in section 4.2. The five lakes are divided into two classes: Northern (GBL, GSL, and LA) and Southern Lakes (SL; LW and LE) as seen in Figure 5.1.

The NLs have a greater mean depth (~20-88m) (Johnson, 1975; Rouse *et al*., 2008; Mitchell & Prepas, 1990), longer residence time (~6-124 years) (Johnson, 1975; Evans, 2000; Gibson *et al*., 2006), and vary in size (7,770-31,153 km^2), compared to the SLs, with LA as the smallest of the selected lakes. NLs are oligotrophic and show limited anthropogenic influence, with mining operations being the most significant pollutant contributor through the

Athabasca River (Bill *et al*., 1994; Schindler, 2001; Timoney, 2008; Parlee & D'Souza, 2019), since there is little to no agriculture/urban land use within the watersheds. Yellowknife Bay of GSL has the largest urban area of the NLs, where an algal bloom did occur in September of 2014 (Pick, 2016). Both GSL and LA exhibit a dimictic mixing classification, whereby vertical lake temperature profiles equalize (~4°C) and allow for epilimnetic and hypolimnetic exchange twice a year (typically during the spring and fall seasons) (Rouse *et al*., 2008; Leconte *et al*., 2008). GBL exhibits a cold monomictic mixing classification, whereby mixing occurs once a year (during summer), therefore only stratifies during the ice cover season. However, recent studies have suggested that the GBL is transitioning from cold monomictic to dimictic for some regions of the lake (Rouse *et al*., 2008).

The SLs are significantly shallower (~12-19m), with shorter residence times (~2.6-5 years) compared to NLs, with a large surface area (23,750-25,744 km^2) (Government of Manitoba, 2021; EPA, 2020). The SLs range in trophic status through different basins, with the south basin of LW and the west basin of LE exhibiting higher algal biomass compared to the north basin of LW and the Central/East basins of LE (Ali & English, 2019; Smith *et al*., 2019). SLs receive external nutrient loading from extensive agricultural development within the surrounding watershed, mainly mobilized by the Saskatchewan and Red River for LW and the Detroit River for LE. Eutrophication due to external nutrient loading has resulted in high algal biomass within predominately the south basin of LW and the west basin of LE. Both LW and the west basin of LE are classified as polymictic, where mixing occurs at irregular intervals, often initiated by high water inflow and wind mobilization, while the deeper Central and East basins of LE are classified as dimictic (Nürnberg & LaZerte, 2016; Karatayev *et al*., 2021). Additionally, both lakes have experienced significant algal blooms, often increasing in intensity/frequency in recent years (Wassenaar *et al*., 2012; Binding *et al*., 2018; Binding *et al*., 2019).
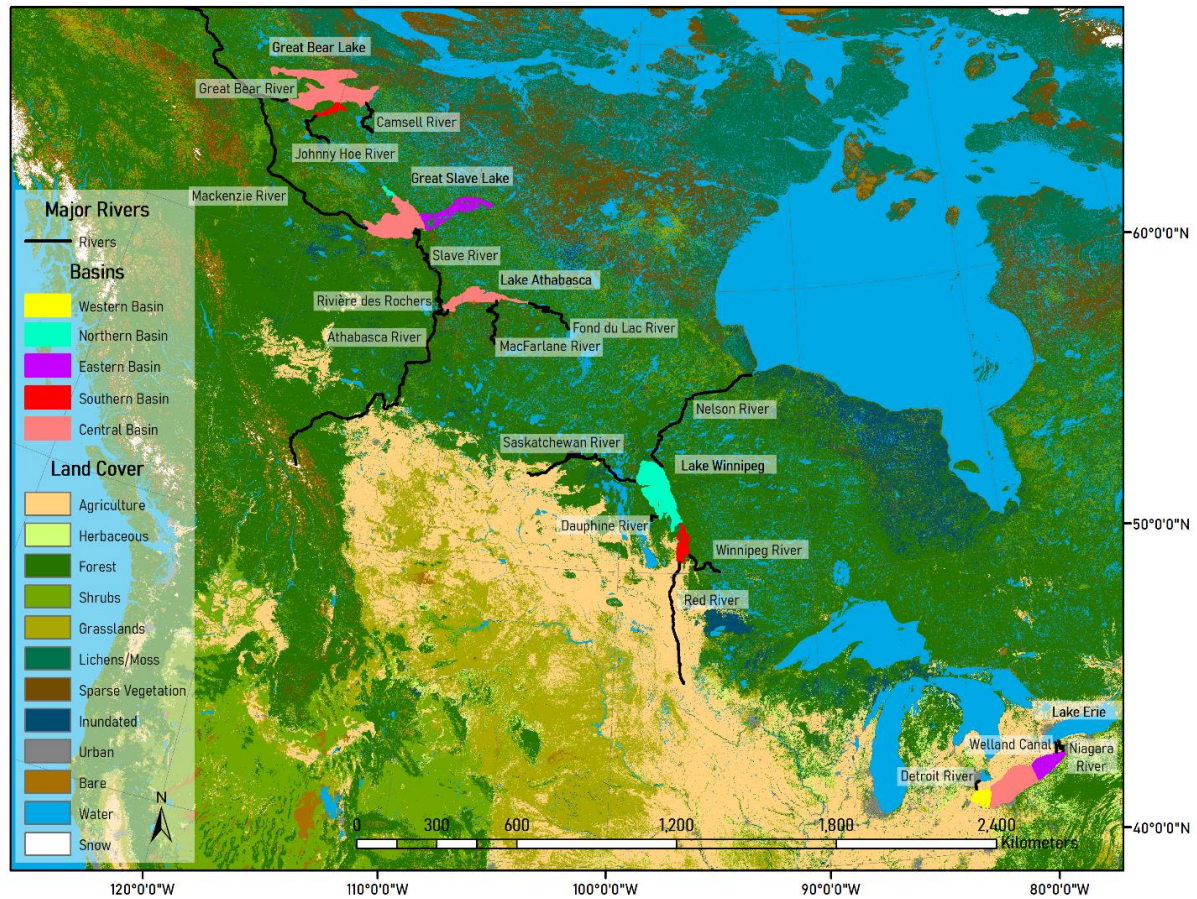
Figure 5.1. Map of the study lakes and their corresponding basins and rivers. Land cover deliniation provided by ESA CCI Landcover. (http://maps.elie.ucl.ac.be/CCI/viewer/download.php).

5.3 Data Sources

This research makes use of the daily remote sensing data product provided by the ESA's CCI Lakes project (Carrea *et al*., 2022). The Lakes project provides a harmonized remote sensing data product on a $1/120^{\circ}$ latitude-longitude grid (approx. 1 km) consisting of various lake properties for over 2,000 lakes globally. Of the lake properties, this research makes use of the LWLR, LIC and LSWT data, which were preprocessed and rescaled to annual and seasonal intervals from 2002-2020. Seasonal intervals were defined by 2-m air temperature data for each lake using the hourly land gridded (0.1° x 0.1°; native resolution of 9 km) climate reanalysis data product (ERA5-Land) provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Preprocessing was performed to derive five annual and seasonal

algal biomass parameters, four annual/seasonal LSWT parameters, and three annual LIC parameters. Algal biomass and LSWT are rescaled annually to allow comparisons to annual changes in the duration and timing of the LIC.

Of the algal biomass parameters, MC and XC represent the average/maximum chl-a concentration per pixel during the ice free (annual) and the temperature defined (spring, summer and fall) seasons. BC represents the average for only chl-a concentrations that exceeded a lake-wide spatial-temporal threshold annually and seasonally. This threshold is seasonally dependent, as, in example, the threshold for spring differs from summer as it only accounts for the historic spring concentrations. BC is indicative of only pixels where chl-a concentrations exceed a z-score of 2, thereby representing only pixels/regions of the lake where chl-a concentrations are relatively greater than the rest of the lake. BC thresholds are lake specific, and therefore use the same threshold for each sub-basin. BD represents the number of days annually/seasonally a chl-a pixel was flagged as abnormally high. These ACCs can be considered a proxy for algal "blooms" since they are either extremely rare, or do not occur in NLs and therefore the thresholds used by other studies (*i.e.*, >10.00 μg L^{-1}), to flag an algal bloom (Binding *et al*., 2021) would not be relevant. As such, these "blooms" are referred to as BC given the available time series. Finally, PC represents the day of year/season a given pixel exhibited the highest chl-a concentration annually/seasonally.

Of the LSWT parameters, MT and XT represent the average/maximum temperature per pixel during the ice-free (annual) and the temperature-defined (spring, summer, and fall) seasons. AT represents the number of days in which the water temperature exceeded a lake wide spatial-temporal threshold annually and seasonally, which is also seasonally and lake dependent. Finally, PT represents the day of year/season for which a given pixel exhibited the highest temperature.

Of the LIC parameters, ON represents the DOY each pixel saw continuous ice cover for a minimum of 14 days, starting per water year (September – August). OF represents the first DOY where each pixel saw continuous open water for a minimum of 14 days. DR

represents the number of days between the ice on and the ice off date per pixel. For a full description of the preprocessing methods, please refer to section 4.4.

5.4 Methods

To determine the capacity of LIC to predict algal biomass concentrations, data were preprocessed at an annual scale, as the minimum temporal resolution of ON, OF and DR is annual. However, LIC may exhibit a more significant control on algal biomass during the spring season compared to later in the year. Therefore, algal biomass parameters were also calculated annually for the spring, summer, and fall seasons. Due to the large size of the study lakes, the lakes were divided into basins, as different regions of a lake may exhibit different water chemistry profiles, depths, and mixing patterns, which therefore may react differently to changes in LIC. To allow for a comparison in predictive performance, LSWT parameters were introduced on the same scale as the algal biomass parameters, as water temperature is known to be a significant controller of algal growth and development (Borowitzka *et al.,* 2016; Cross *et al.,* 2015). This research assesses predictive performance in two different ways: (1) prediction of annual/seasonal algal biomass parameters via regressors, and (2) prediction of annual/seasonal abnormal algal biomass flags (presence of a BC pixel) via classifiers. Regressors assessed in this study are linear (multiple linear regression; MLR) and non-linear (feed-forward artificial neural network; ANN).

To predict annual and seasonal algal biomass parameters (MC, XC, BC, BD and PC), an MLR was calculated for each lake/basin and season, using only LIC parameters ON, OF, DR), only LSWT (MT, XT, AT, PT) parameters and both parameters. An MLR is calculated as linear combination of each independent input parameter via the following equation (1):

$$Y = \alpha + X_1 \cdot \beta_1 + X_2 \cdot \beta_2 + \cdots + X_n \cdot \beta_n \qquad (5.1)$$

Where Y is the predicted algal biomass parameter, α is the bias, $X_{1 \cdots n}$ is the input parameters, and $\beta_{1 \cdots n}$ is the weights. The chl-*a* concentration data were log scaled for the MLR to increase the likelihood of a normal distribution, as chl-*a* concentrations are typically left skewed. A stratified random sampling method is employed to select training data for chl-*a*

concentration measured algal biomass parameters (MC, XC and BC), where a random 250 pixels for each threshold of chl-*a* concentrations (< 2.5 µg L⁻¹, 2.5 – 5.0 µg L⁻¹, 5.0 – 10.0 µg L⁻¹, 10.0 – 20.0 µg L⁻¹, > 20.0 µg L⁻¹) are selected for each year (2002 – 2011 and 2016 – 2020; excluding 2012-2015 due to missing data). Stratified random sampling is used to ensure a wide range of chl-*a* concentrations to train the model, with a maximum sample size of 18,750 independent replicates. The sample size will vary by lake, season, and year, as some may not have 250 observations above 20.0 µg L⁻¹, and some NLs may see ice cover during the spring season, limiting the number of available observations. For both BD and PC, the training data are selected randomly, with 1,000 pixels selected each year/season, comprising a total of 15,000 observations for each lake/basin. To validate the predicted chl-*a* concentrations, a 10-fold cross-validation approach is implemented, resulting in a 90% / 10% training/testing data split, ensuring that observations not included in the model can be accurately predicted. The resulting MLRs and predicted algal biomass parameters are used to provide various performance metrics; coefficient of determination (r^2), adjusted coefficient of determination (adj. r^2), p-value, standard error (SE), Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (NRMSE), Median Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE), as described in table 5.1.

Table 5.1. Summary of performance and error metrics for the MLR and ANN (adj. r2 and Se are limited to the MLR only)

Parameter	Equation	Variables	Measure	#
Adj. r^2	$1 - \left[\dfrac{(1 - r^2) \times (n - 1)}{(n - k - l)}\right]$	n = sample size, r^2 = coefficient of determination, k = number of independent variables	Corrected (adjusted) coefficient of determination (r^2) for determining model accuracy	(5.2)
S_e	$\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{df}}$	n = sample size, y_i = the observed value, $\hat{y}_i$ = the predicted value, df = degrees of freedom	Standard error of the residuals indicating model fit and variance	(5.3)
RMSE	$RMSE = \sqrt{\sum_{i=1}^{n}\dfrac{(y_i - \hat{y}_i)^2}{n}}$	n = sample size, y_i = the observed value, $\hat{y}_i$ = the predicted value	Standard deviation of the prediction errors	(5.4)
NRMSE	$NRMSE = \dfrac{RMSE}{\sigma}$	$RMSE$ = Root Mean Square Error, σ = standard deviation of input dependent data	RMSE normalized by the standard deviation of the input training data. Allows for comparison between models of varying values	(5.5)

Parameter	Equation	Variables	Measure	#
MAE	$MAE = \dfrac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n}$	n = sample size, y_i = the observed value, $\hat{y}_i$ = the predicted value	Average error between pairs of predicted and observed values	(5.6)
MAPE	$MAPE = 100 \times median\ of\ \left[\dfrac{|\hat{y}_i - y_i|}{y_i}\right] for\ i = 1, \dots, n$	n = sample size, y_i = the observed value, $\hat{y}_i$ = the predicted value	Median of the absolute percentage errors between the predicted and observed values	(5.7)

Each error metric is calculated per fold and averaged to get a mean error across all folds. While RMSE is a standard error metric, the study area contains lakes with a wide range of algal biomass concentrations, where GBL is highly oligotrophic, while LW is eutrophic. Therefore, as an example, an RMSE of 4.00 µg L^{-1} indicates a much poorer performance for GBL compared to LW. The normalization of the RMSE by the standard deviation of the observed values allows for better model performance comparison between different lakes.

The interaction of LIC, LSWT and algal biomass may not exhibit a significant linear trend, where instead non-linear learners may provide a better predictive performance. Machine learning allows for the identification of complex patterns between dependent and independent variables. This study implements a feed-forward artificial neural network (ANN) for predicting each algal biomass parameter using only LIC, LSWT and both inputs (training data selection the same as for the MLR). The ANN was constructed using Keras for R and Tensorflow (Abadi *et al.,* 2016; Falbel *et al.,* 2019), and has two hidden layers with 64 and 32 nodes. As input training data are measured at differing scales (*i.e.*, days, K, DOY), they were first normalized to the same scale to ensure proper convergence and stabilize the training process via the following equation:

$$z = \frac{X - \mu}{\sigma} \tag{5.8}$$

Where X = the input parameter, μ = the mean and σ = the standard deviation of the input parameters. Using the input data, the activation of each node is defined by the scaled exponential linear units (SELU) activation function, which is described by the following equation (Klambauer *et al.,* 2017):

$$selu(x) = \lambda \begin{cases} x & if \ x > 0 \\ \alpha e^x - a & if \ x \le 0 \end{cases} \tag{5.9}$$

Where α = 1.6733 and λ = 1.0507. If the given value of each node is greater than zero, the node is scaled by λ, while inputs with a value of less than 0 are scaled with an additional constant. While the ReLU activation function is more common, there are issues where the neurons will consistently be calculated as zero, known as the dying ReLU problem. Dying ReLUs tend to occur when the learning rate is too high, or there are large negative biases, which is more likely to occur when the same ANN structure is applied to varying datasets (*i.e.*, differing algal biomass parameters and lakes). The SELU activation function allows for node values of less than zero, and self normalizes, improving model training time. To calculate the global minima, the Adam optimization function was used with a batch size of 32, a learning rate of 0.001, a total of 100 epochs, and an internal training/validation split of 90%/10% where mean squared error was used as the internal error metric. Adam optimization was selected as it is computationally efficient, applicable for models with a high number of observations and suited for noisy gradients (Kingma & Ba, 2015). An early callback was implemented to reduce over-fitting, where if the internal mean squared does not decrease after 10 epochs the training was ended. The same ANN structure was trained using annual/seasonal input data for each lake/basin and season to predict for each algal biomass parameter. Performance metrics (RMSE, NRMSE, MAE and MAPE) were used to calculate each ANN performance, to allow for comparisons between lakes/seasons and determine whether models constructed from only LIC or LSWT parameters provided greater predictive performance. ANN may provide evidence for the performance of LIC and LSWT in predicting algal biomass concentrations, however, this does not provide evidence for its predictive capacity of abnormal agal biomass conditions or potential "bloom flags". Instead, a classifier is constructed using a Random Forest (RF).

To determine the capacity of LIC and LSWT to classify abnormal algal biomass conditions, an RF model is constructed using annual/seasonal LIC and LSWT parameters. The BC product was converted to a binary integer where all cells with a value were converted to

the value 1 flag (indicating abnormal biomass), and all NA pixels within the lake boundary were converted to the value 0 flag (indicating no abnormal biomass). This conversion is done to create a classifier capable of identifying the presence of relatively (lake specific) high chl-a concentrations based solely on the LIC and LSWT parameters. For each year of data (excluding 2012-2015 due to missing data), a stratified random sampling of 1,000 abnormal and 1,000 no abnormal flagged pixels were extracted, with their corresponding LIC and LSWT pixel values. The RF model was constructed using all annual/seasonal LIC and LSWT parameters for each lake/basin, where a maximum of 100 trees were constructed with 80% / 20% training and testing data split, with a total of 10 RF iterations. The RF model was constructed using the 'randomForest' package in r (Liaw & Wiener, 2002), which was ported from Breiman (2001).

While the RF model can provide evidence for accuracy in classifying abnormally high algal biomass parameters in each lake/season, it is also important to determine which input parameters had greater classification importance. The internal model error loss (variable importance) is calculated using an out-of-bag approach, where the model error is calculated per tree when each input parameter is excluded, with a mean standard error calculated across all trees. The 'randomForest' package calculates the internal model error loss and subsequently provides the mean decrease accuracy (MDA) across all trees when each parameter is excluded. The MDA is scaled by the standard deviation of the decrease in accuracy across the trees to improve the stability of the results (Breiman, 2001). The parameter with the largest MDA represents the parameter with the greatest importance. Comparisons of where and when the parameter importance changes have led to an understanding of the role LIC plays in classifying abnormal algal biomass in North American Great Lakes.

5.5 Results

The MLR method provided relatively poor performances compared to the ANN for all parameters of algal biomass, with the exception of the PC parameter (Figure 5.2). The use of LSWT parameters showed a significant improvement in the predictions of algal biomass parameters compared to models that used only LIC.
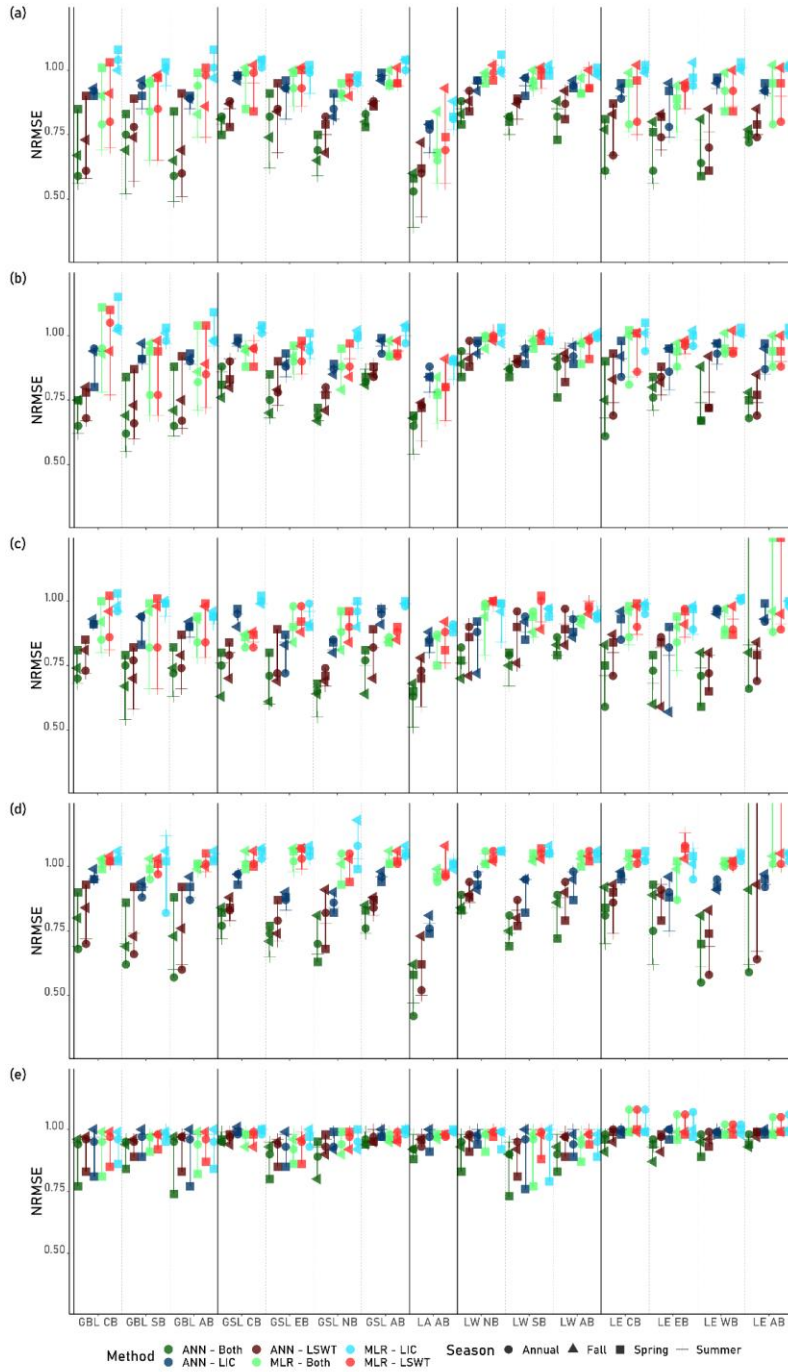
Figure 5.2. Normalized Root Mean Squared Error (NRMSE) of the Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) for five algal biomass parameters: (a) mean chl-*a* concentrations, (b) maximum chl-*a* concentrations, (c) abnormal chl-*a* concentrations, (d) abnormal chl-*a* days, and € peak chl-*a* DOY. Where: GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, LE = Lake Erie, CB = Central Basin, SB = South Basin, NB = North Basin, EB = East Basin, WB = West Basin, and AB = All Basins. Data are separated into models that contain only lake surface water temperature parameters (LSWT), only lake ice cover parameters (LIC), or contain both LSWT and LIC parameters.

110

5.5.1 Multiple Linear Regression

Across all seasons and lakes, LSWT only models produced a median NRMSE of 0.95 compared to 1.00 produced by LIC only models for MC, with 0.95 and 1.01 for XC, 0.91 and 0.98 for BC, 1.03 and 1.05 for BD, and 0.98 and 0.99 for PC. The inclusion of both LIC and LSWT parameters typically reduced predictive error, with a median NRMSE of 0.94, 0.94, 0.88 1.01 and 0.98 for MC, XC, BC, BD and PC, respectively (Table 5.2).

Across all lakes and model inputs, median NRMSE was lowest in the summer season for MC (0.93), XC (0.94) and BC (0.89) compared to annual, (0.95, 0.95, 0.91), spring (0.97, 0.98, 0.96), and fall (1.00, 0.98, 0.95) seasons. Annual data returned the lowest median NRMSE for predicting BD (1.02) compared to spring (1.02), summer (1.03) and fall (1.04), while spring returned the lowest mean NRMSE for PC (0.97), compared to annual (0.98), summer (0.99) and fall (0.99).

Across all seasons and model inputs, LA returned the lowest median NRMSE when predicting MC, XC, BC, and BD (0.76, 0.82, 0.88 & 0.99 respectively), while GBL (0.96, 0.97, 0.94 & 1.02 respectively), GSL (0.96, 0.95, 0.90 & 1.03 respectively), LW (0.99, 0.99, 0.96 & 1.05 respectively) and LE (0.95, 0.97, 0.95 & 1.02 respectively) performed significantly poorer on average. Predictions for PC returned GBL with the lowest, albeit poor, median NRMSE (0.97) compared to GSL, LA, LW and LE (0.98, 0.98, 0.98 & 1.00 respectively).

The NRMSE varies between lake basins, seasons, and model inputs, where the best performances occurred when models contained both LSWT and LIC inputs during the summer season for all algal biomass parameters (median NRMSE = 0.56 – 1.04) (Figure 5.2). Of the models using both LIC and LSWT parameters in the summer season, LA returned the lowest median NRMSE for MC, XC, BC, and BD (0.56, 0.67, 0.76, & 0.99 respectively), while GBL returned the lowest median NRME for peak chl-*a* DOY (0.98). Within this same grouping, the NLs (GBL, GSL and LA) returned a lower median NRMSE for all algal biomass parameters (0.80 – 1.02, median = 0.94) compared to the SL (LW and LE) (0.90 – 1.04, median = 0.99), particularly for chl-*a* concentration parameters (median NRMSE of NLs = 0.83, SLs = 0.93). Of the NLs, for chl-*a* concentration parameters, all basins of GBL and LA had a relatively

lower median NRMSE (0.68 – 0.77), while all of GSL returned a relatively high NRMSE (0.89 – 0.94). Of the SLs, LW returned the poorest results for all algal biomass parameters with little difference between the basins (0.98 – 1.04). The performance for LE returns the lowest NRMSE for the Central basin for all chl-*a* concentration parameters (0.75 – 0.87) compared to the other basins (0.85 – 0.95).

Table 5.2. Lake basin range of mean seasonal MLR performance metrics where NRMSE = normalized root mean square error, RMSE = root mean square error (µg L^{-1} for mean/max/abnormal chl-*a*, days for abnormal chl-*a* days and peak chl-*a* DOY), MAPE = median absolute percentage error, MAE = mean absolute error (µg L^{-1} for mean/max/abnormal chl-*a*, days for abnormal chl-*a* days and peak chl-*a* DOY), Adj. R2 = adjusted coefficient of determination, and se = standard error (µg L^{-1} for mean/max/abnormal chl-*a*, days for abnormal chl-*a* days and peak chl-*a* DOY)

		Lake	NRMSE	RMSE	MAPE	MAE	Adj. r^2	se
Mean Chl-*a*	LSWT	GBL	0.86 - 0.90	2.01 - 2.73	0.30 - 0.53	1.33 - 1.69	0.34 - 0.51	0.48 - 0.85
		GSL	0.94 - 0.96	2.17 - 2.61	0.29 - 0.32	1.55 - 1.82	0.09 - 0.18	0.43 - 0.54
		LA	0.73	3.24	0.30	2.22	0.45	0.46
		LW	0.97 - 1.00	3.75 - 4.26	0.26 - 0.36	2.89 - 3.25	0.07 - 0.12	0.43 - 0.54
		LE	0.88 - 0.92	1.95 - 3.59	0.29 - 0.43	1.19 - 1.95	0.23 - 0.31	0.45 - 0.59
	LIC	GBL	1.00 - 1.04	2.29 - 3.07	0.39 - 0.72	1.63 - 2.05	0.07 - 0.15	0.57 - 1.11
		GSL	0.96 - 1.03	2.23 - 2.80	0.30 - 0.31	1.59 - 1.86	0.01 - 0.12	0.44 - 0.57
		LA	0.82	3.67	0.34	2.56	0.33	0.52
		LW	0.99 - 1.01	3.83 - 4.32	0.27 - 0.38	2.93 - 3.30	0.04 - 0.06	0.44 - 0.54
		LE	0.97 - 1.02	1.97 - 4.05	0.33 - 0.45	1.25 - 2.23	0.03 - 0.17	0.52 - 0.64
	Both	GBL	0.85 - 0.88	1.99 - 2.69	0.29 - 0.52	1.32 - 1.67	0.35 - 0.53	0.47 - 0.84
		GSL	0.91 - 0.96	2.16 - 2.59	0.28 - 0.33	1.54 - 1.81	0.10 - 0.22	0.42 - 0.54
		LA	0.68	3.04	0.28	2.1	0.51	0.44
		LW	0.94 - 0.98	3.68 - 4.19	0.25 - 0.36	2.81 - 3.17	0.10 - 0.15	0.42 - 0.53
		LE	0.86 - 0.92	1.78 - 3.54	0.28 - 0.38	1.09 - 1.92	0.28 - 0.35	0.44 - 0.55
Max Chl-*a*	LSWT	GBL	0.84 - 0.97	2.88 - 3.98	0.34 - 0.52	1.98 - 2.32	0.34 - 0.46	0.51 - 0.84
		GSL	0.90 - 0.94	3.60 - 3.92	0.29 - 0.32	2.45 - 2.92	0.14 - 0.22	0.44 - 0.54
		LA	0.80	3.97	0.30	2.71	0.37	0.45
		LW	0.97 - 1.00	5.71 - 6.62	0.24 - 0.33	3.99 - 4.14	0.06 - 0.11	0.39 - 0.50
		LE	0.93 - 0.96	3.01 - 9.23	0.31 - 0.40	1.69 - 3.13	0.18 - 0.27	0.47 - 0.59
	LIC	GBL	1.00 - 1.06	3.41 - 4.42	0.44 - 0.68	2.45 - 2.78	0.06 - 0.13	0.61 - 1.04
		GSL	0.96 - 1.02	3.77 - 4.20	0.30 - 0.34	2.54 - 3.18	0.03 - 0.13	0.47 - 0.57
		LA	0.88	4.39	0.34	3.03	0.26	0.49

	Lake	NRMSE	RMSE	MAPE	MAE	Adj. r^2	se
	LW	0.99 - 1.00	5.72 - 6.69	0.24 - 0.34	3.96 - 4.26	0.03 - 0.06	0.39 - 0.51
	LE	0.98 - 1.01	3.09 - 8.36	0.35 - 0.42	1.76 - 3.42	0.03 - 0.12	0.52 - 0.66
Both	GBL	0.84 - 0.94	2.86 - 3.95	0.34 - 0.52	1.96 - 2.29	0.35 - 0.47	0.51 - 0.83
	GSL	0.88 - 0.94	3.55 - 3.89	0.28 - 0.32	2.42 - 2.85	0.14 - 0.24	0.43 - 0.54
	LA	0.76	3.78	0.29	2.59	0.43	0.43
	LW	0.96 - 0.99	5.59 - 6.57	0.24 - 0.32	3.85 - 4.09	0.08 - 0.14	0.38 - 0.49
	LE	0.92 - 0.96	2.89 - 7.84	0.30 - 0.38	1.58 - 3.09	0.22 - 0.31	0.46 - 0.58
Abnormal Chl-*a* — LSWT	GBL	0.87 - 0.91	2.71 - 3.04	0.28 - 0.40	1.59 - 2.10	0.28 - 0.32	0.47 - 0.62
	GSL	0.85 - 0.92	1.96 - 3.36	0.16 - 0.21	1.61 - 2.03	0.20 - 0.30	0.21 - 0.28
	LA	0.84	3.42	0.22	2.67	0.32	0.31
	LW	0.96 - 1.00	4.21 - 7.32	0.12 - 0.16	2.81 - 3.71	0.02 - 0.13	0.22 - 0.26
	LE	0.92 - 0.94	1.71 - 4.55	0.16 - 0.25	1.19 - 2.29	0.13 - 0.24	0.25 - 0.35
LIC	GBL	0.94 - 0.98	2.80 - 3.18	0.35 - 0.46	1.81 - 2.25	0.10 - 0.20	0.53 - 0.68
	GSL	0.94 - 1.00	2.11 - 3.39	0.18 - 0.26	1.75 - 2.30	0.04 - 0.18	0.23 - 0.31
	LA	0.89	3.70	0.24	2.91	0.23	0.33
	LW	0.94 - 0.96	4.23 - 7.00	0.14 - 0.15	2.92 - 3.44	0.09 - 0.13	0.22 - 0.25
	LE	0.93 - 1.01	1.77 - 5.00	0.17 - 0.30	1.25 - 2.60	0.01 - 0.20	0.26 - 0.40
Both	GBL	0.86 - 0.89	2.62 - 2.98	0.28 - 0.38	1.58 - 2.04	0.31 - 0.37	0.46 - 0.60
	GSL	0.85 - 0.90	1.93 - 3.28	0.16 - 0.20	1.58 - 1.97	0.22 - 0.31	0.21 - 0.27
	LA	0.81	3.30	0.21	2.55	0.37	0.30
	LW	0.91 - 0.92	3.99 - 6.89	0.12 - 0.14	2.68 - 3.31	0.13 - 0.21	0.20 - 0.24
	LE	0.85 - 0.92	1.67 - 4.55	0.16 - 0.24	1.16 - 2.26	0.18 - 0.30	0.24 - 0.35
Abnormal Chl-*a* Days — LSWT	GBL	1.01 - 1.02	7.66 - 21.30	0.51 - 0.61	4.34 - 16.82	0.06 - 0.16	0.84 - 0.94
	GSL	1.01 - 1.04	6.97 - 11.13	0.51 - 0.61	4.10 - 7.16	0.06 - 0.14	0.93 - 0.95
	LA	1.00	9.12	0.61	5.00	0.10	0.96
	LW	1.04 - 1.05	6.63 - 6.74	0.58 - 0.61	4.46 - 4.62	0.04 - 0.05	0.91 - 0.94
	LE	1.01 - 1.08	9.80 - 39.51	0.60 - 0.76	4.76 - 24.15	0.03 - 0.08	0.96 - 1.31
LIC	GBL	1.01 - 1.04	7.84 - 21.09	0.54 - 0.62	4.41 - 18.16	0.02 - 0.08	0.89 - 0.95
	GSL	1.05 - 1.07	7.08 - 11.37	0.50 - 0.62	4.20 - 7.17	0.01 - 0.09	0.96 - 0.98
	LA	1.01	9.26	0.62	5.04	0.08	0.97
	LW	1.05 - 1.06	6.69 - 6.78	0.58 - 0.62	4.50 - 4.70	0.01 - 0.03	0.93 - 0.95
	LE	1.02 - 1.04	9.87 - 29.35	0.60 - 0.71	4.78 - 17.36	0.01 - 0.20	0.97 - 1.20

	Lake	NRMSE	RMSE	MAPE	MAE	Adj. r^2	se
Both	GBL	1.00 - 1.01	7.62 - 21.37	0.50 - 0.60	4.30 - 16.72	0.07 - 0.17	0.84 - 0.93
	GSL	1.01 - 1.03	6.93 - 10.91	0.51 - 0.61	4.08 - 7.02	0.07 - 0.18	0.93 - 0.94
	LA	0.97	8.89	0.60	4.85	0.14	0.94
	LW	1.03 - 1.04	6.58 - 6.66	0.57 - 0.60	4.41 - 4.60	0.05 - 0.07	0.91 - 0.93
	LE	0.97 - 1.02	9.76 - 27.48	0.59 - 0.68	4.73 - 16.29	0.04 - 0.26	0.96 - 1.15
LSWT	GBL	0.94 - 0.96	18.53 - 19.60	0.07 - 0.07	15.57 - 16.29	0.08 - 0.11	0.09 - 0.09
	GSL	0.94 - 0.97	20.45 - 22.96	0.06 - 0.08	16.13 - 19.43	0.05 - 0.12	0.10 - 0.11
	LA	0.98	26.28	0.10	22.1	0.04	0.14
	LW	0.96 - 0.99	23.04 - 27.43	0.06 - 0.11	18.28 - 23.65	0.03 - 0.09	0.11 - 0.14
	LE	1.00 - 1.02	35.31 - 45.19	0.13 - 0.18	29.93 - 39.90	0.01 - 0.03	0.22 - 0.30
LIC	GBL	0.94 - 0.98	18.45 - 19.89	0.07 - 0.07	15.89 - 16.88	0.05 - 0.11	0.09 - 0.09
	GSL	0.97 - 1.00	20.70 - 22.73	0.07 - 0.08	16.44 - 19.21	0.01 - 0.08	0.10 - 0.11
	LA	0.99	26.73	0.10	22.81	0.02	0.14
	LW	0.94 - 0.98	22.93 - 27.14	0.06 - 0.11	18.06 - 23.38	0.05 - 0.12	0.11 - 0.13
	LE	1.00 - 1.02	35.48 - 45.22	0.13 - 0.19	30.18 - 40.01	0.01 - 0.02	0.22 - 0.30
Both	GBL	0.92 - 0.96	18.09 - 19.46	0.07 - 0.07	15.34 - 16.15	0.10 - 0.14	0.09 - 0.09
	GSL	0.93 - 0.97	20.38 - 22.50	0.06 - 0.08	16.12 - 18.98	0.06 - 0.13	0.10 - 0.10
	LA	0.97	26.14	0.10	22.00	0.06	0.13
	LW	0.93 - 0.97	22.72 - 26.76	0.06 - 0.11	17.93 - 22.76	0.08 - 0.15	0.11 - 0.13
	LE	0.99 - 1.02	35.13 - 45.10	0.13 - 0.18	29.68 - 39.75	0.02 - 0.05	0.22 - 0.30

(Row group label spanning left: Peak Chl-a DOY)

5.5.2 Artificial Neural Network

The ANN method provided improved performance compared to the MLR for all algal biomass parameters, with the exception of the PC parameter (ANN median NRMSE = 0.81, MLR median NRMSE = 0.95). Similar to the MLR, the use of LSWT parameters provided improvements to the prediction of algal biomass parameters, compared to models using only LIC.

Across all seasons and lakes, LSWT only models produced a median NRMSE of 0.82 compared to 0.93 produced by LIC only models for MC, with 0.80 and 0.93 for XC, 0.79 and 0.91 for BC, 0.83 and 0.94 for BD, and 0.96 and 0.98 for PC. The inclusion of both LIC and

LSWT parameters typically reduced predictive error, with a median NRMSE of 0.75, 0.76, 0.74, 0.75 and 0.94 for MC, XC, BC, BD and PC, respectively (table 5.3).

Across all lakes and model inputs, median NRMSE was lowest in the summer season for MC (0.76), XC (0.84), AC (0.79) and BD (0.78) compared to annual (0.82, 0.84, 0.79 & 0.78), spring (0.85, 0.87, 0.84 & 0.87), and fall (0.83, 0.85, 0.79 & 0.89) seasons. Spring data returned the lowest median NRMSE for predicting PC (0.91), compared to annual (0.96), summer (0.98) and fall (0.96).

Across all seasons and model inputs, LA returned the lowest median NRMSE when predicting MC, XC, BC, and BD (0.61, 0.73, 0.72 & 0.62 respectively), while GBL (0.81, 0.77, 0.80 & 0.85 respectively), GSL (0.84, 0.85, 0.80 & 0.83 respectively), LW (0.88, 0.91, 0.86 & 0.88 respectively) and LE (0.79, 0.83, 0.81 & 0.90 respectively) performed poorer on average. Predictions for PC returned LW with the lowest, albeit poor, median NRMSE (0.95) compared to GBL, GSL, LA and LE (0.96, 0.96, 0.96 & 0.98 respectively).

Similar to the MLR, the ANN indicated the best performance for models that contained both LSWT and LIC inputs during the summer season for all algal biomass parameters (median NRMSE = 0.43 – 0.98) (Figure 5.2). Of the models that used LIC and LSWT in the summer season, LA returned the lowest median NRMSE for MC, XC, BC and BD (0.43, 0.59, 0.59, & 0.50 respectively), while GBL returned the lowest median NRME for PC (0.96). Within this same grouping, NLs (GBL, GSL and LA) returned a lower median NRMSE for all algal biomass parameters (0.63 – 0.97, median = 0.74) compared to SLs (LW and LE) (0.76 – 0.99, median = 0.83), particularly for chl-*a* concentration parameters (median NRMSE of NLs = 0.68, SLs = 0.80). Of the NLs, for the chl-*a* concentration parameters, all the GBL and LA basins had a relatively low median NRMSE (0.54 – 0.69), while the GSL Central basin returned the highest NRMSE (0.80 – 0.91). Of the GSL basins, the Eastern and Northern basins returned the lowest median NRMSE for all chl-*a* concentration parameters (0.69 – 0.76). Of the SLs, LW returned the poorest results for all algal biomass parameters with little difference between the basins (0.86 – 0.99). However, the Southern basin of LW provided considerably lower NRMSE (0.75) compared to the Northern basin (0.91) for ACC. The performance of LE varies

115

depending on the parameter, with MC returning lower NRMSE for the Central (0.67) and Eastern (0.69) basins compared to the Western basin (0.76), while BC returns lower NRMSE for the Western basin (0.69) compared to the Central (0.80) and (0.82) Eastern basins. Based on the findings presented, LIC acts as a poor predictor of any algal biomass parameter as LSWT consistently provides lower predictive error, irrespective of seasonality, location, or method (linear vs. non-linear). While LIC may not provide a strong prediction of a given algal biomass parameter value, it may aid in the classification of lake-wide BC flags.

Table 5.3. Lake basin range of mean seasonal ANN performance metrics where NRMSE = normalized root mean square error, RMSE = root mean square error (μg L^{-1} for mean/max/abnormal chl-*a*, days for abnormal chl-*a* days and peak chl-*a* DOY), MAPE = median absolute percentage error, MAE = mean absolute error (μg L^{-1} for mean/max/abnormal chl-*a*, days for abnormal chl-*a* days and peak chl-*a* DOY), Adj. R2 = adjusted coefficient of determination, and se = standard error (μg L^{-1} for mean/max/abnormal chl-*a*, days for abnormal chl-*a* days and peak chl-*a* DOY)

		Lake	NRMSE	RMSE	MAPE	MAE
Mean Chl-*a*	**LSWT**	GBL	0.68 - 0.74	1.74 - 2.21	0.27 - 0.50	1.20 - 1.57
		GSL	0.76 - 0.87	1.87 - 2.36	0.23 - 0.30	1.37 - 1.73
		LA	0.59	2.6	0.25	1.84
		LW	0.86 - 0.89	3.31 - 3.82	0.22 - 0.32	2.54 - 2.93
		LE	0.73 - 0.78	1.51 - 2.97	0.25 - 0.36	1.01 - 1.70
	LIC	GBL	0.88 - 0.92	2.09 - 2.64	0.37 - 0.75	1.55 - 1.97
		GSL	0.85 - 0.98	2.07 - 2.67	0.28 - 0.32	1.53 - 1.85
		LA	0.76	3.31	0.34	2.44
		LW	0.94 - 0.95	3.60 - 4.08	0.26 - 0.36	2.83 - 3.23
		LE	0.85 - 0.95	1.73 - 3.88	0.33 - 0.42	1.14 - 2.33
	Both	GBL	0.64 - 0.70	1.64 - 2.10	0.26 - 0.50	1.14 - 1.47
		GSL	0.67 - 0.81	1.64 - 2.19	0.19 - 0.28	1.24 - 1.60
		LA	0.52	2.31	0.22	1.63
		LW	0.80 - 0.84	3.04 - 3.60	0.20 - 0.29	2.31 - 2.70
		LE	0.68 - 0.74	1.38 - 2.82	0.23 - 0.31	0.85 - 1.63
Max Chl-*a*	**LSWT**	GBL	0.72 - 0.74	2.44 - 3.45	0.30 - 0.51	1.73 - 2.23
		GSL	0.76 - 0.86	3.18 - 3.61	0.24 - 0.32	2.21 - 2.63
		LA	0.7	3.44	0.27	2.4
		LW	0.90 - 0.92	5.26 - 6.32	0.22 - 0.31	3.64 - 4.01
		LE	0.76 - 0.83	2.53 - 6.56	0.30 - 0.38	1.55 - 3.03

	Lake	NRMSE	RMSE	MAPE	MAE
LIC	GBL	0.91 - 0.92	3.13 - 4.01	0.42 - 0.66	2.40 - 2.73
	GSL	0.87 - 0.98	3.50 - 4.02	0.30 - 0.38	2.49 - 2.93
	LA	0.84	4.12	0.35	3.01
	LW	0.93 - 0.95	5.44 - 6.47	0.22 - 0.34	3.77 - 4.22
	LE	0.89 - 0.94	2.87 - 7.91	0.39 - 0.43	1.68 - 3.67
Both	GBL	0.68 - 0.71	2.30 - 3.30	0.28 - 0.49	1.60 - 2.08
	GSL	0.69 - 0.84	2.86 - 3.50	0.20 - 0.30	2.03 - 2.53
	LA	0.64	3.16	0.25	2.22
	LW	0.87 - 0.90	5.06 - 6.18	0.20 - 0.29	3.45 - 3.83
	LE	0.74 - 0.78	2.34 - 6.16	0.29 - 0.35	1.41 - 2.91
LSWT	GBL	0.72 - 0.78	2.35 - 2.59	0.26 - 0.35	1.44 - 1.81
	GSL	0.70 - 0.81	1.53 - 3.06	0.11 - 0.18	1.20 - 1.84
	LA	0.7	2.89	0.16	2.13
	LW	0.84 - 0.90	3.71 - 6.93	0.11 - 0.15	2.50 - 3.57
	LE	0.74 - 0.80	1.46 - 4.28	0.13 - 0.20	1.02 - 1.91
LIC	GBL	0.89 - 0.92	2.64 - 2.96	0.36 - 0.44	1.80 - 2.19
	GSL	0.81 - 0.95	1.79 - 3.21	0.14 - 0.24	1.44 - 2.27
	LA	0.84	3.47	0.22	2.74
	LW	0.84 - 0.90	3.99 - 6.86	0.12 - 0.15	2.77 - 3.46
	LE	0.77 - 0.96	1.67 - 4.86	0.17 - 0.32	1.23 - 2.71
Both	GBL	0.69 - 0.73	2.22 - 2.45	0.25 - 0.32	1.37 - 1.69
	GSL	0.63 - 0.74	1.37 - 2.84	0.10 - 0.16	1.06 - 1.68
	LA	0.62	2.54	0.14	1.86
	LW	0.75 - 0.82	3.31 - 6.36	0.10 - 0.13	2.22 - 3.13
	LE	0.70 - 3.30	1.31 - 11.98	0.12 - 0.18	0.91 - 1.92
LSWT	GBL	0.72 - 0.80	8.75 - 14.12	0.21 - 0.60	6.36 - 10.94
	GSL	0.78 - 0.85	10.01 - 13.71	0.34 - 0.65	7.40 - 10.15
	LA	0.59	13.61	0.43	10.32
	LW	0.82 - 0.89	10.72 - 11.44	0.54 - 0.62	7.94 - 8.40
	LE	0.71 - 1.63	14.29 - 32.45	0.41 - 0.58	10.12 - 25.92
LIC	GBL	0.91 - 0.96	10.94 - 19.09	0.26 - 0.69	8.37 - 16.34
	GSL	0.86 - 0.96	11.27 - 15.76	0.34 - 0.84	8.63 - 12.28

Abnormal Chl-a (rows 1–30) and *Abnormal Chl-a Days* (rows 31–32) are the major row groupings on the left margin.

		Lake	NRMSE	RMSE	MAPE	MAE
Peak Chl-*a* DOY	**Both**	LA	0.77	18.91	0.52	15.1
		LW	0.92 - 0.95	11.52 - 12.72	0.62 - 0.68	8.78 - 9.25
		LE	0.87 - 0.96	16.51 - 28.64	0.55 - 0.66	11.86 - 22.05
		GBL	0.70 - 0.77	8.20 - 13.67	0.20 - 0.58	5.83 - 10.44
		GSL	0.70 - 0.79	9.27 - 12.64	0.29 - 0.63	6.80 - 9.28
		LA	0.52	11.73	0.37	8.73
		LW	0.75 - 0.84	10.20 - 10.83	0.51 - 0.58	7.22 - 7.88
		LE	0.67 - 1.89	13.53 - 35.39	0.39 - 0.54	9.56 - 17.66
	LSWT	GBL	0.93 - 0.94	18.26 - 19.22	0.06 - 0.06	15.13 - 15.74
		GSL	0.92 - 0.97	20.32 - 22.18	0.06 - 0.07	16.10 - 18.41
		LA	0.96	25.74	0.09	21.49
		LW	0.91 - 0.96	22.63 - 26.26	0.06 - 0.10	17.72 - 22.10
		LE	0.96 - 0.98	34.12 - 43.27	0.11 - 0.16	27.52 - 38.02
	LIC	GBL	0.93 - 0.96	18.29 - 19.57	0.06 - 0.07	15.34 - 16.09
		GSL	0.94 - 1.00	20.68 - 22.49	0.07 - 0.07	16.60 - 18.79
		LA	0.97	26.31	0.09	22.46
		LW	0.93 - 0.97	22.77 - 26.80	0.06 - 0.10	17.78 - 22.99
		LE	0.99 - 0.99	34.74 - 43.44	0.12 - 0.17	28.39 - 38.30
	Both	GBL	0.90 - 0.92	17.79 - 18.94	0.06 - 0.06	14.62 - 15.41
		GSL	0.89 - 0.96	20.02 - 21.10	0.06 - 0.07	15.88 - 17.20
		LA	0.92	25.23	0.09	20.91
		LW	0.87 - 0.92	21.93 - 25.13	0.06 - 0.09	17.08 - 20.80
		LE	0.92 - 0.96	33.56 - 42.45	0.11 - 0.16	26.90 - 36.59

5.5.3 Random Forest Classifier

The RF model returned variable predictive accuracy depending on the lake and season (72.88% – 97.27%, median = 84.22%) (Figure 5.3). When sorted by seasonality, BC flags in the summer season had the highest overall accuracy across all lakes (76.89% – 97.27%, median = 88.56%), followed by annual, spring and fall data (Table 5.4).
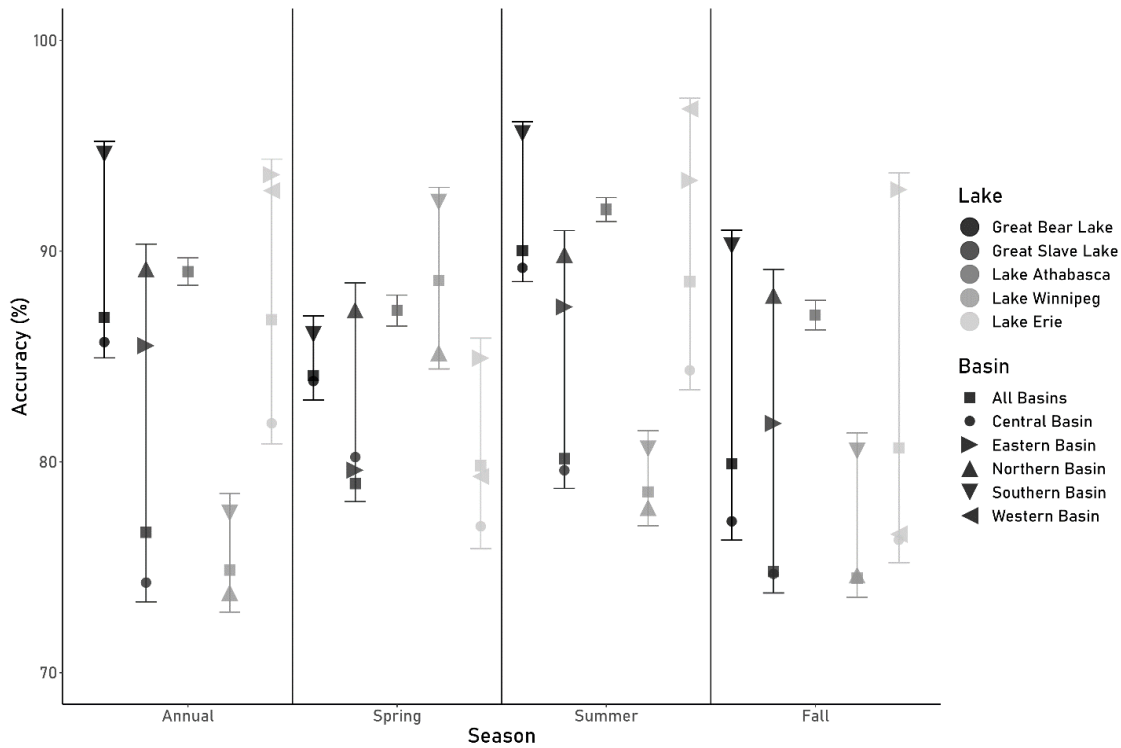
Figure 5.3. Mean random forest (RF) classification accuracy of the presence of abnormal algal biomass per lake, basin, and season across 10 RF model iterations. Error bars represent the minimum and maximum lower and upper boundary tree accuracy.

Table 5.4. Random Forest classification accuracy of the presence of abnormal algal biomass where data is summarized per season across all lakes where Min. = minimum lower boundary accuracy, Max. = maximum upper boundary accuracy, Med. = median accuracy of all lakes per season, Pos. Pred. Value = median accuracy for positive classifications, Neg. Pred. Value = median accuracy for negative classifications, and P = percentage of lakes where the accuracy p-value is < 0.05.

	Min.	Max.	Med.	Pos. Pred. Value	Neg. Pred. Value	Precision	P
Annual	72.88%	95.21%	85.68%	85.47%	81.02%	85.47%	100%
Spring	75.89%	93.03%	84.10%	81.44%	85.78%	81.44%	100%
Summer	76.96%	97.27%	88.56%	86.66%	84.44%	86.66%	100%
Fall	73.57%	93.71%	79.91%	77.70%	78.29%	77.70%	100%

While most lakes show higher classification accuracy in the annual and summer models, LW indicates higher classification accuracy during the spring season compared to any other season (84.41 – 93.03%, median = 88.61%) (Figure 5.3). LA had the highest overall accuracy across all seasons (86.26 – 92.54%, median = 88.11%) followed by GBL, LE, GSL, and LW (Table 5.5). The NLs returned a higher accuracy (73.36% – 96.15%, median = 85.87%) compared to the SLs (72.88% – 97.27%, median = 79.62%).

119

Table 5.5. Random Forest classification accuracy of the presence of abnormal algal biomass where data are summarized per lake across all seasons, where Min. = minimum lower boundary accuracy, Max. = maximum upper boundary accuracy, Med. = median accuracy of all seasons per lake, Pos. Pred. Value = median accuracy for positive classifications, Neg. Pred. Value = median accuracy for negative classifications, and P = percentage of lakes where accuracy p-value is < 0.05.

	Min.	Max.	Med.	Pos. Pred. Value	Neg. Pred. Value	Precision	P
GBL	76.29%	96.15%	86.46%	86.20%	86.27%	86.20%	100%
GSL	73.36%	90.97%	80.19%	79.69%	80.38%	79.69%	100%
LA	86.26%	92.54%	88.11%	88.97%	87.64%	88.97%	100%
LW	72.88%	93.03%	78.21%	78.55%	77.45%	78.55%	100%
LE	75.22%	97.27%	84.63%	82.18%	81.85%	82.18%	100%

Within each lake, the SB of GBL had the highest accuracy (85.15% - 96.15%, median = 92.42%) in BC classification, with the NB of GSL (85.85% – 90.97%, median = 88.53%), the SB of LW (76.62% – 93.03%, median = 80.56%) and the EB of LE (83.94% – 94.37%, median = 93.13%) having the corresponding highest accuracy (Table 5.6). The difference in accurate positive to accurate negative classifications remained within ±5.00% for 71.67% (43/60) of RF models for varying lakes and seasons, and has a median difference of ±1.08%. The 28.33% (17/60) RF models with a positive to negative accuracy difference exceeding ±5.00% returned a median difference of ±11.24%. This difference occurred in 17 RF models, of which 41.18% (7/17) occurred in LE, 29.41% (5/17) in GBL, 23.53% (4/17) in GSL, 5.88% (1/17) in LW, and none in LA, with a near even split between seasons. The precision and accuracy of BC classification remained within ±5.00% for 91.67% of RF models (median difference = ±0.74%). Of the 8.33% (5/60) RF models where the difference in accuracy and precision exceeded ±5.00% (median difference = ±7.32%), three occurred in the Eastern basin of LE (for three seasons: annual, summer and fall), once in the Central basin of GBL (spring), and once in the Eastern basin of GSL (spring).

Table 5.6. Random Forest classification accuracy where data are summarized per lake basin across all seasons, where Min. = minimum lower boundary accuracy, Max. = maximum upper boundary accuracy, Med. = median accuracy of all seasons per lake basin, Pos. Pred. Value = median accuracy for positive classifications, Neg. Pred. Value = median accuracy for negative classifications, and P = percentage of lakes where accuracy p-value is < 0.05.

Lake	Min.	Max.	Med.	Pos. Pred. Value	Neg. Pred. Value	Precision	P
GBL Central Basin	76.29%	89.84%	84.76%	81.59%	86.98%	81.59%	100%

Lake	Min.	Max.	Med.	Pos. Pred. Value	Neg. Pred. Value	Precision	P
GBL Southern Basin	85.15%	96.15%	92.42%	93.70%	83.16%	93.70%	100%
GBL All Basins	79.07%	90.64%	85.48%	83.70%	86.72%	83.70%	100%
GSL Central Basin	73.36%	81.05%	77.14%	77.08%	77.21%	77.08%	100%
GSL Eastern Basin	78.63%	88.09%	83.66%	82.35%	84.84%	82.35%	100%
GSL Northern Basin	85.85%	90.97%	88.53%	90.91%	83.25%	90.91%	100%
GSL All Basins	73.89%	80.98%	77.82%	77.33%	78.34%	77.33%	100%
LA All Basins	86.26%	92.54%	88.11%	88.97%	87.64%	88.97%	100%
LW Northern Basin	72.88%	85.91%	76.23%	75.76%	76.71%	75.76%	100%
LW Southern Basin	76.62%	93.03%	80.56%	79.32%	79.28%	79.32%	100%
LW All Basins	73.57%	89.26%	76.73%	76.50%	76.96%	76.50%	100%
LE Central Basin	75.22%	85.23%	79.38%	80.05%	79.32%	80.05%	100%
LE Eastern Basin	83.94%	94.37%	93.13%	83.87%	94.59%	83.87%	100%
LE Western Basin	75.52%	97.27%	86.08%	88.31%	76.22%	88.31%	100%
LE All Basins	78.90%	89.28%	83.72%	83.62%	84.39%	83.62%	100%

To determine variable importance in classifying BC presence, the MDA was calculated across 10 RF iterations of different training and testing data. The distribution of the MDA across all RF iterations for each LSWT and LIC input parameter per lake and season is shown in Figure 5.4. The magnitude of the MDA is relative to the input data and may not be directly comparable to other RF models; therefore, it is the ranking of each variable that is important. Each parameter is ranked from 1 (greatest mean accuracy decrease) to 7 (lowest mean accuracy decrease) for each RF iteration and averaged to determine the overall ranking of variable importance as seen in Figure 5.5. Across all lakes, annual RF models showed higher importance for LSWT parameters compared to LIC, where AT returned the highest mean importance ranking (3.09), while OF (3.89) was ranked fifth. Across all lakes, spring RF models returned PT as the highest ranking (2.49) on average; however, OF jumps to second rank (3.29) indicating increased importance in the timing of ice melt during spring. The importance of LIC drops in the later seasons, as across all lakes in summer, PT is ranked first

(2.66) and XT during fall (2.48), while OF drops to third (3.61) and fourth (3.66) respectively (Table 5.7).
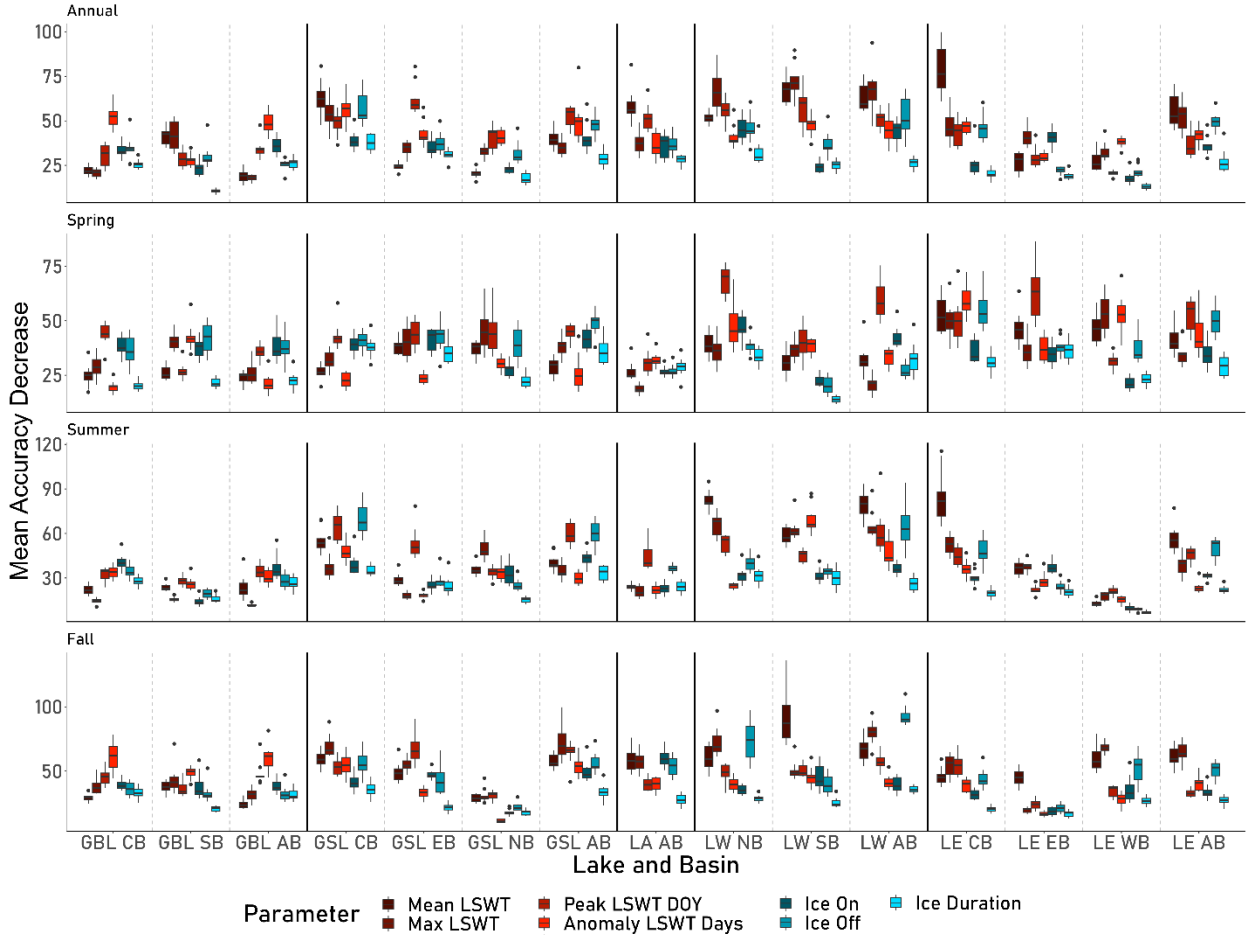
Figure 5.4. Boxplot of Mean Decrease Accuracy (MDA) distributions of Random Forest (RF) classification models for abnormal algal biomass presence per lake, basin and season across 10 RF model iterations. Black dots represent outliers, and horizontal lines represent the median MDA. Where: GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, LE = Lake Erie, CB = Central Basin, SB = South Basin, NB = North Basin, EB = East Basin, WB = West Basin, and AB = All Basins.

Table 5.7. Random Forest Mean Decrease Accuracy median rankings across all lakes per season, where MT = Mean Lake Surface Water Temperature (LSWT), XT = Maximum LSWT, PT = Peak LSWT DOY, AT = Anomaly LSWT Days, ON = Ice On date, OF = Ice Off date, and DR = Ice Duration.

	MT	XT	PT	AT	ON	OF	DR
Annual	3.39	3.31	3.31	3.09	4.61	3.89	6.41
Spring	4.31	4.23	2.49	4.12	4.00	3.29	5.56
Summer	2.98	4.04	2.66	4.37	4.47	3.61	5.87
Fall	3.11	2.48	3.30	4.35	4.69	3.66	6.41

Across all seasons in GBL, AT returned the highest ranking (2.78), while ON and OF returned third and fourth, respectively (3.23 & 3.69). Across all seasons, both GSL and LA returned PT as the highest ranking (2.35 & 2.80) on average, with OF ranked second and third respectively (3.09 & 3.40). Across all seasons, both LW and LE returned the same mean rank orders, with MT as the highest ranking (2.63 & 2.44), while OF is the most important LIC variable at fourth (4.13 & 3.74) (Table 5.8).

Table 5.8. Random Forest Mean Decrease Accuracy median rankings across all seasons per lake, where MT = Mean Lake Surface Water Temperature (LSWT), XT = Maximum LSWT, PT = Peak LSWT DOY, AT = Anomaly LSWT Days, ON = Ice On date, OF = Ice Off date, and DR = Ice Duration.

	MT	XT	PT	AT	ON	OF	DR
GBL	4.98	4.77	2.93	2.78	3.23	3.69	5.63
GSL	3.95	3.49	2.35	4.79	4.46	3.09	5.87
LA	3.28	4.88	2.80	4.23	4.08	3.40	5.35
LW	2.63	2.78	2.90	4.22	4.97	4.13	6.38
LE	2.44	2.81	3.59	3.84	5.04	3.74	6.52

Based on mean ranking summaries, LIC parameters show greater importance for NLs and during the spring season. When rankings are averaged by both lake and season, LIC parameter importance increases during spring for high latitude lakes, as both GBL and GSL show OF as the highest-ranking variable (2.33 & 2.33), while DR is ranked third (3.40). In comparison LA during spring ranks DR as the third highest (3.40), ON third highest in LW (3.37), and OF as fourth highest in LE (3.33). Spring months typically show high LIC importance, often ranked first or second behind PT, with more LIC parameters at a higher ranking in NLs, with some variability (*e.g*., low LIC importance in the Northern basin of GSL, high OF importance in the Central basin of LE) (Figure 5.5). LIC importance drops later in the year, along with PT, as MT and XT return higher importance in the summer and fall (most notably for LW and LE).
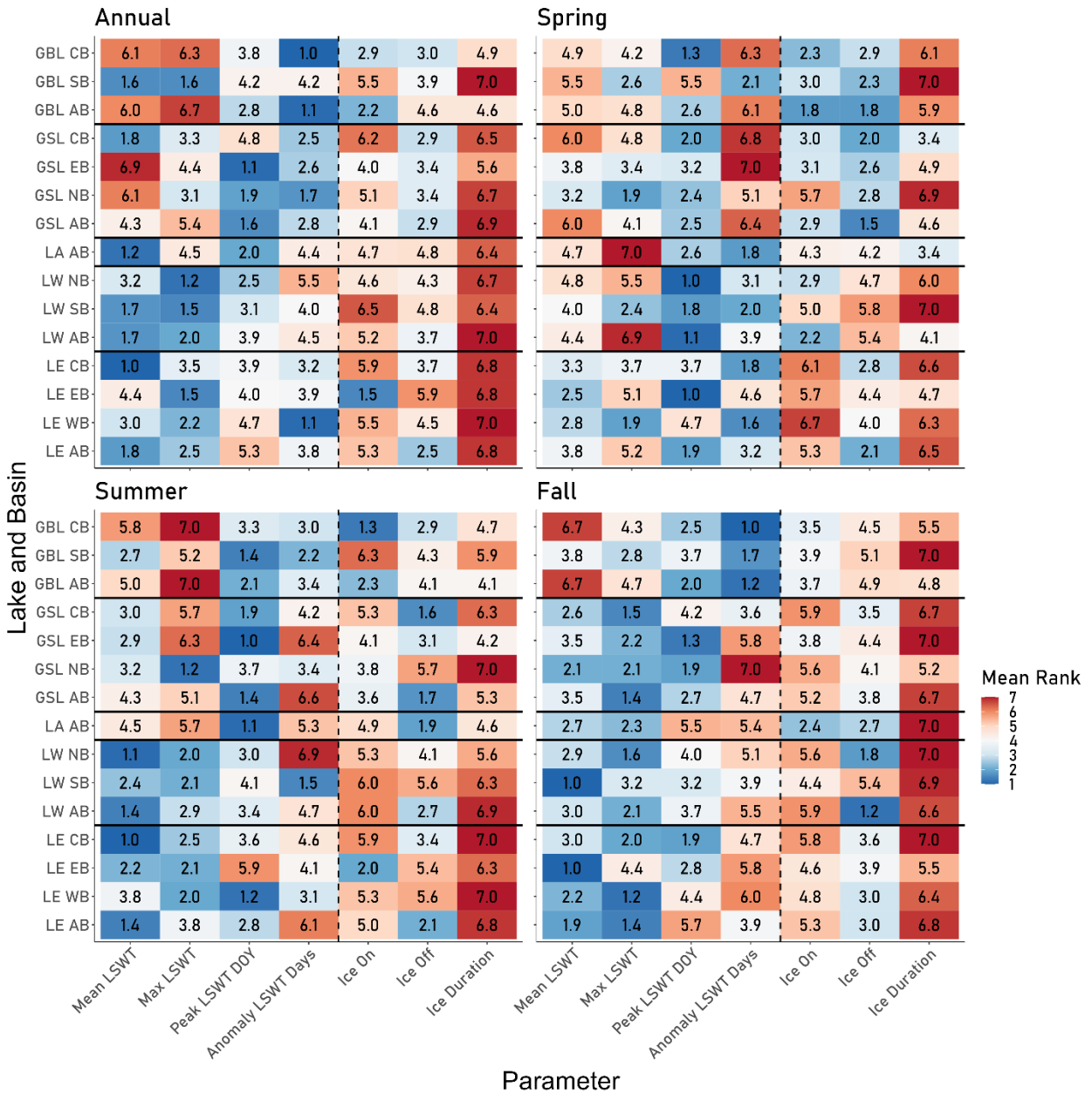
Figure 5.5. Heat Map representing the mean ranking of the Mean Decrease Accuracy (MDA) variable importance for a Random Forest (RF) classifier of abnormal algal biomass presence across lakes, basins, and seasons. The lower the value (blue), the greater the variable importance on average across 10 RF model iterations. Where: GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, LE = Lake Erie, CB = Central Basin, SB = South Basin, NB = North Basin, EB = East Basin, WB = West Basin, and AB = All Basins.

5.6 Discussion

5.6.1 Prediction of Algal Biomass Parameters

Between the two regression methods, the ANN consistently outperforms the MLR, returning lower error metrics with the exception of PC, as both methods returned poor results. This is expected as machine learning methods typically provide improved performance over linear regressors (Chen *et al.,* 2019b; Amemiya *et al.,* 2023). The BD returned the greatest difference between the ANN and MLR performance. Machine learning can capture non-linear patterns, in which the improved performance of the ANN indicates the relationship between algal biomass, LSWT and LIC are non-linear. The ANN may also be further improved by adapting the internal network structure to each lake, season, and algal biomass parameter, rather than generalized, as was done in this research. In particular, the PC parameter may have returned improved performance utilizing an alternate ANN structure, due to the poor performance of PC when using the same structure for all algal biomass parameters. However, the PC input data had substantial variance, as the chl-*a* peak was calculated as the DOY when chl-*a* was highest. Therefore, the same pixel position could return high chl-*a* in spring one year and fall the next year, resulting in highly variable data. Improvements to interpolation, and spatial/temporal smoothing could have reduced the PC variance and improved predictive capacity; however, lake surface chl-*a* peaks can still occur at any point in the ice-free period, depending on atmospheric, lake physical, and biogeochemical variables (Zhang *et al.,* 2011; Tian *et al.,* 2017). The PC parameter did not improve depending on the predictor parameters, as the inclusion of either LSWT or LIC did not show significant differences in performance.

The inclusion of both LSWT and LIC into the same model returned a lower error for both the MLR and the ANN, despite the poor performance of the LIC only models. Ice duration was consistently found to be the least important variable across all lakes, basins, and seasons for the RF. It is possible that the inclusion of ice duration in the LIC only MLR and ANN models increases the predictive error. Additional ANN and MLR models are needed to determine whether individual LIC parameters have improved predictive capacity over LSWT parameters. A total of 120 combinations of LSWT and LIC parameters would be required,

resulting in a total of 36,000 models (15 lakes/basins, 4 seasons, and 5 algal biomass parameters). Due to the training time required for the ANN, the data were only separated into LSWT only, LIC only, or both. Some studies have shown that the increase in inputs for machine learning models return improved performance (Trofimova *et al.,* 2021), however, this is not guaranteed (Arif *et al.,* 2020; Qiu *et al.,* 2023; Domingos, 2012). It is possible that the LIC parameters do improve chl-*a* concentration predictions when included with LSWT. The inclusion of LIC parameters may have improved chl-*a* concentration predictions when combined with LSWT parameters, as the ratio in the timing between OF and MT/XT was a significant contributor to the algal blooms observed in the remote and traditionally oligotrophic Dickson Lake in 2014 (Favot *et al.,* 2019). Previous research on the remote and traditionally oligotrophic Lake Dickson (Algonquin, Ontario, Canada) had found that the ratio in the timing between OF and MT/XT was a significant contributor to the 2014 algal bloom (Favot *et al.,* 2019). The timing between lake ice melt and stratification onset reduced the duration of the mixing season, where the lake did not mix fully, resulting in lower hypolimnetic oxygen (Favot *et al.,* 2019). Low oxygen increases internal loading, where low redox potential breaks Fe-P complexes, increasing P availability (Tammeorg *et al.,* 2020; Harrow-Lyle & Kirkwood, 2021). The timing between ice melt and stratification may not be as impactful for the deep waters of the NLs, however, shallow basins within each lake (*i.e*. GBL SB and GSL NB) may be affected in such a manner. LIC parameters alone do not provide a strong predictive accuracy of algal biomass parameters as LIC is indicative of only the winter conditions, while LSWT is measured at the same time as the algal biomass parameters. The combination of winter conditions and LSWT in the same model provides a marked improvement in algal biomass predictions, particularly during the summer.

Predictive accuracy greatly increased for chl-*a* concentration predictions during the summer months, particularly for NLs. The prevalence of lake ice is still high during the spring season for lakes such as GBL and GSL, limiting the number of available satellite observations for open water. Additionally, misclassification between ice and water, and increased observable reflectance due to melting, broken ice in the water may have resulted in greater

over prediction of chl-*a* concentrations during the spring season. Lake ice formation may also occur during the fall, resulting in similar issues with spring chl-*a* retrieval. Additionally, cloud coverage for the region is much higher during this time, limiting observations, thus affecting retrieval accuracy (Konik *et al.,* 2019). The spring and fall seasons are also considerably shorter than the summer, and therefore limit the number of available observations in which to derive a seasonal mean, increasing variance compared to the summer season. For these reasons, both annual and summer derived data return a lower predictive error. The contribution of lake ice during the summer may be related to stratification, increasing internal loading later in the year due to a lower redox potential (Woolway *et al.,* 2021). An improved summer performance was also observed in LE within the more oligotrophic and mesotrophic Eastern and Central basins for MC and XC. Meanwhile, in the west basin of LE, all chl-*a* concentration parameters returned the highest predictive accuracy during spring, possibly because summer and fall conditions are more closely related to external nutrient loading. Lake depth may play a role in the seasonal predictive accuracy of the BC. Shallower basins (relative to the rest of the lake) such as the south basin of GBL, the north basin of GSL, and the south basin of LW returned greater differences in summer predictive accuracy to other seasons. Shallow waters stratify quickly, leading to earlier anoxia driven internal loading (Woolway *et al.,* 2021), however, are prone to frequent mixing via precipitation driven flushing and high wind speeds, increasing turnover during the summer and subsequent epilimnetic available P (Robertson & Diebel, 2020). Conversely, deeper waters such as the Central basin of GBL and GSL, the Eastern basins of GSL and LE, and the north basin of LW returned a stronger predictive performance of BC during the fall season. Deep waters take longer to stratify (Boehrer & Schultze, 2008), where epilimnetic P increases through turnover and diffusion (Robertson & Diebel, 2020; Tammeorg *et al.,* 2020), resulting in increased observable surface algae during the fall. Therefore, the summer months for deep lakes may not experience a strong response in summer algae to high LSWTs, and more strongly affect the algae observed in the fall due to prolonged stratification.

Regions with high external loading (such as the SLs) during the growing season due to anthropogenic activities (herein referred to as anthropogenic loading) may introduce variance where the LSWT and LIC parameters will not fully explain the variance seen in the algal biomass response. The NLs typically outperformed the SLs in all seasons. Deep water regions of the NLs are often nutrient poor, where there is limited anthropogenic loading, and therefore the majority of nutrients are derived via internal loading (Finney *et al.,* 2004). External loading is also high in some regions of the NLs, such as the CB of GSL, where the Slave River discharges various nutrients (Evans & Muir, 2016). Such loading however, is not dependent on anthropogenic activities, and therefore driven by natural systems. The NLs are therefore reliant on LSWT and LIC for both the timing of mixing, the extent/severity of hypolemntic anoxic diffusion, and river discharge. These NLs are also deeper than LW and LE, where diffusion would be more prevalent. Although NLs performed well, LA returned the best performance for all algal biomass parameters, with the exception of PC, with both ANN and MLR. LA is the shallowest of the three NLs, is oligotrophic, and has minimal anthropogenic influence with the exception of mining and oil sands operations runoff. There have been reports of increased algal growth by indigenous communities, primarily attributed to discharge from oil sands operations (Bill *et al.,* 1994; Timoney, 2008; Parlee & D'Souza, 2019). The oil sands have been shown to increase total polycyclic aromatic hydrocarbons (PAHs) into waters (Smol, 2019), which studies have found to increase algal growth (Anderson *et al.,* 2010; Harris & Smith, 2016). There has been an observed increase in estimated PAH mass deposition in the Athabasca Oil Sands region in recent years (Chibwe *et al.,* 2021; Lévesque *et al.,* 2023), the origin of which may be from natural bitumen deposits (Hall *et al.,* 2012). It is possible that the increase in PAHs in combination with warming waters returned stronger predictive models, as it has been observed that climate warming has reduced the sedimentation burial of PAHs (Tao *et al.,* 2019). However, the LIC only models for LA returned low predictive accuracy compared to the LSWT, indicating a similar performance to the other lakes. There is little evidence to support that LIC parameters provide a strong predictor of algal biomass parameters, regardless of season, lake, or basin. LIC only models returned a similar error across all lakes, which is

consistently lower than the LSWT only parameters. However, its inclusion with LSWT parameters significantly improves predictive accuracy.

5.6.2 Classification of Abnormal Algal Biomass Presence

Classification returned adequate accuracy for most lakes on average, with fall typically lower for most lake basins and highest in summer. Spring indicates the lowest variability in RF model performance (Figure 5.3), however, less accurate than the annual and summer seasons. This is interesting considering the high variance of the spring input data and the poor performance of the ANN and MLR models. Additionally, the Central basin of GSL and all of LW show the highest accuracy in spring compared to any other season. The summer and fall season in LW would likely show improved accuracy with the inclusion of biogeochemical variables (*e.g.*, P, N, DO, etc.) and even lake mixing information, given LW classified as polymictic (Nürnberg & LaZerte, 2016). The importance of OF and PT is likely a proxy of spring and snow melt, coinciding with internal/external loading. The same should also be true for the Western basin of LE where the accuracy was highest in spring compared to the other basins but still poorer in comparison to summer. The odd pattern of LW performance may also stem from data product issues, as see in section4.5. Interestingly, regardless of season, relatively shallower and more eutrophic lake basins show higher precision and accuracy compared to deeper and more oligotrophic regions of the lake (Table 5.6), such as the south basin of GBL, the north basin of GSL, the south basin of LW, and the west basin of LE (east basin has a higher median accuracy but lower precision). Shallower waters have a stronger response to changes in temperature (Havens *et al.,* 2016) and experience faster ice loss (Li *et al.,* 2022). Additionally, all of these basins (with the exception of the NB of GSL), intersect with major river inflows, where nutrients are most frequently mobilized. The external nutrient concentrations are not dependent on anthropogenic activities, therefore the LSWT and LIC parameters provide stronger predictive capacity for the NLs (due to their control on nutrient availability timings).

Of the LIC parameters used, ON and OF were among the most important variables for predicting abnormal algae during the spring season, where OF was typically more important

129

for NLs. The timing of ice melt corresponds to both the onset of mixing and internal loading via turbulence (Vincent *et al.,* 2012). Additionally, an early OF period extends the growing season and typically increases stratification duration due to prolonged solar radiative flux exposure (Dibike *et al.,* 2011). The importance of OF was also associated with the highly important PT, in which an earlier timing may result in a more rapid onset of stratification (Woolway *et al.,* 2021). LIC parameters may not be considered important for SLs due to frequent ice beak-up, and periodic snow melt during the winter season, particularly for regions such as the Western basin of LE and the south basin of LW. While ON and OF were found to be important, DR was not, returning the lowest ranking on average. There were difficulties in defining ON dates due to high cloud cover, which would increase the variability in the DR calculation, affecting the discernible relation between DR and algal biomass. The duration of ice cover may not be important, given that nutrient and light availability can stimulate algal growth primarily when the OF period begins. The trend of later ON formation may not result in higher spring algal biomass formation, as a study by Hébert *et al*. (2021) found that later ice formation increased the survivability of zooplankton overwintering, initiating stronger top-down controls during spring. Therefore, this complicates the relationship of DR to the algal biomass response, as later ON may reduce algal biomass; however, earlier OF may facilitate higher growth, likely resulting in the low importance of DR found in the RF. The importance of LIC parameters decreases in the later season on average (Figure 5.5). However, LA and LW still show high importance for OF during summer and fall, potentially due to prologued stratification induced by an earlier OF date. The RF models provide evidence that the LIC parameters are important in the classification of annual/seasonal abnormal algal biomass during the spring for NLs, along with the timing of algal biomass peaks.

5.7 Conclusion

Despite the importance of LIC, it was not found to be a significant predictor of algal biomass concentrations but did improve linear and non-linear predictive models when included with LSWT parameters. A combination of the timing/duration of LIC and the measure of LSWT provides the best models for chl-a concentration prediction. The timing of ON, and in

particular OF, was of significant importance in classifying abnormal algal biomass conditions in spring, primarily for NLs. The application of LIC to algal biomass change is typically not explored, often due to data limitations, assumptions of ecological dormancy during winter, and frequently blooming/commonly studied lakes are either not ice-forming (*e.g.*, Lake Taihu, Lake Okeechobee), or experience limited/intermittent ice cover (*e.g.*, Lake Erie). This research has shown that LIC is an important variable in the springtime algal biomass response for NLs, which are experiencing LIC reduction at an accelerated rate. The inclusion of LIC parameters in algal biomass projection models may be important for ice-forming lakes. The reduction of LIC is a consequence of climate change, indicating that climate-driven factors are important predictors of algal biomass trends. NL models showed greater performance compared to SLs, potentially indicating a greater control of climatic variables on algal biomass dynamics for NLs. This research was limited to large lakes, where smaller lakes may exhibit a stronger response to environmental change. This research was additionally limited to annual scales, where other parameters (*e.g.*, wind speed, precipitation) would likely have a significant intra-annual impact on the algal biomass response. Future research may look to apply the given methods to other study regions or may look at the impact of other climatic parameters as they affect algal biomass within a shorter timeframe. Algal biomass dynamics may be better explained by additional lake physical and atmospheric parameters. Such research can improve our understanding of the role atmospheric and lake physical parameters play on the dynamics of algae, and how they differ between lakes.

Chapter 6: The Role of Lake Physical Variables and Atmospheric Forcings on the Change in Algal Biomass in North American Great Lakes

6.1 Introduction

Traditional understanding of freshwater algal biomass dynamics is attributed primarily to lake trophic conditions, where macronutrients (*e.g.*, Phosphorus (P), Nitrogen (N), Carbon (C), etc.) and micronutrients (*e.g.*, Iron (Fe), Copper (Cu), etc.) play a significant role in facilitating algal growth (Schindler, 1974; Bonachela *et al.,* 2011; Song *et al.,* 2012; Zhang *et al.,* 2019). Unprecedented algal growth for a given lake or lake basin can result in a bloom, or bloom-like conditions. There has been a perceived increase in the frequency, severity, and duration of algal blooms globally (Favot *et al.,* 2023; Winter *et al.,* 2011; Paerl & Huisman, 2008). Although the external anthropogenic load of nutrients (agricultural or urban runoff/wastewater) is an important driver, blooms have been observed in traditionally oligotrophic lakes, with minimal anthropogenic influence (Favot *et al.,* 2019; Pick, 2016; Winter *et al.,* 2011; Smol, 2019). Climate change has been hypothesized to play a significant role in recent algal bloom trends, primarily due to the increase in global air temperatures (Paerl & Huisman, 2008). It is often assumed that an increase in air temperature will result in higher water temperatures, where most algae exhibit increased growth rates (Borowitzka *et al.,* 2016; Cross *et al.,* 2014; Gobler, 2020). The interaction between water and air temperatures are more complex, however, where various atmospheric forcings (*e.g.*, precipitation, windspeed, net solar radiation, etc.) and lake physical variables (*e.g.*, morphometry, salinity, turbidity, lake flow rate, lake mixing depth/duration/timing, etc.) can affect not only water temperatures but also algal biomass dynamics (Rose *et al.,* 2016; Walsh *et al.,* 2020). The complexity of these interactions often limits our understanding of how climate change may drive not only algal biomass dynamics, but also bloom conditions.

Global water temperatures have been observed to increase by ~0.30°C per decade (Komatsu *et al.,* 2007; O'Reilly *et al.,* 2015; Yang *et al.,* 2019), with regional studies observing increases between 0.23-0.46°C per decade (Filatov *et al.,* 2019; Niedrist *et al.,* 2018; Roberts

et al., 2017; Woolway *et al.,* 2019), with a higher rate of change at higher latitudes (Zhang *et al.,* 2019). Most algae exhibit a higher growth rate at increased water temperatures; however, this trend is not linear, and growth rates decline past an optima (Singh & Singh, 2015; Paerl & Otten, 2013). Cyanobacteria typically have the highest growth temperature optima and are therefore anticipated to dominate algae community compositions in warmer waters (Paerl & Otten, 2013).

Prolonged high-water temperatures with increased lake water column stability (low inflow, precipitation, and windspeed) and high net solar radiation, often results in stratification. Stratified waters limit nutrient exchange between the epilimnion and hypolimnion, where epilimnetic waters are warmer and often associated with higher algal biomass. The epilimnion during prolonged stratification periods can become nutrient poor, while in the hypolimnion, the decomposition of decaying algae results in anoxic conditions. Anoxic waters increase internal loading and, therefore, the availability of P, where mobile algae (*e.g.*, flagellates and cyanobacteria) have advantageous fixation in stratified waters. Mobile algae may travel vertically within the water column to assimilate nutrients and return to the photic zone for favorable growth conditions (higher light availability, warmer waters) within a limited range (Brasil *et al.,* 2016; Lehman *et al.,* 2017).

Conversely, when a lake mixes, the water column becomes oxygenated and nutrients are mobilized via internal loading (Orihel *et al.,* 2017). The turbulence of the water during mixing typically minimizes algal growth due to turbulence; however, mixing events followed by stratification are often associated with higher algal growth (Salmaso, 2010). The onset of mixing varies between lakes, where some mix only once a year (monomictic), twice a year (dimictic), or frequently (continuous and discontinuous polymictic), often driven by changes in water temperature, windspeed, precipitation, and runoff. Studies have found that short mixing events (in which the lake does not fully oxygenate) exhibited prolonged anoxic conditions, leading to increased internal loading and subsequent algal blooms, even in oligotrophic systems (Favot *et al.,* 2019).

133

The majority of research on the impacts of atmospheric forcings and lake physical variables on algal biomass is often limited to either controlled lab studies (Hennon & Dyhrman, 2020; Griffith & Gobler, 2020), mesocosm experiments (Trochine *et al.,* 2010; Moss *et al.,* 2003; 2004), or small-scale observations (Vilhena *et al.,* 2010; Qin *et al.,* 2021). The availability of water quality data often limits the potential scope of interaction-based analysis. Recent advances in remote sensing and computing technologies have allowed the widespread availability of comprehensive water quality data products. This research uses the European Space Agency (ESA), Climate Change Initiative (CCI) Lakes product (ver. 2.0.0) (Carrea *et al.,* 2022), which provides daily surface water chlorophyll-*a* concentrations (chl-*a*; a proxy of algal biomass) and lake surface water temperatures (LSWT), for five North American Great Lakes (Great Bear Lake, Great Slave Lake, Lake Athabasca, Lake Winnipeg, and Lake Erie) and their basins. Additional atmospheric and lake physical parameters are derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-Land data climate reanalysis product, where daily mean 2-m air temperature (T2m), 10-m east and north wind speed (WS), lake mixing level depth (LMLD), surface net solar radiation (SNSR), and daily total precipitation (PPT), surface runoff (SR), and subsurface runoff (SSR) were used (Muñoz-Sabater *et al.,* 2021).

This research seeks to define the interaction effects of atmospheric and lake physical variables on algal biomass dynamics (2002-2020) utilizing a High Order Dynamic Gaussian Bayesian Network (HO-DGBN). The HO-DGBN must first provide adequate chl-*a* forecasting accuracies, ensuring that the network and input parameters provide a significant explanation of the algal biomass dynamics. To determine the interaction effects, this research looks to (1) determine the frequency and combination of significant parameters within temporal models across all lakes, (2) determine the nature of these interactions, (3) determine the importance of input parameters, and (4) asses the sensitivity of the models to changes in atmospheric and lake physical variables. A DGBN will be created utilizing not only a complete time series, but also the stationary, non-stationary and residuals of the training data inputs. The stationary, non-stationary, and error signals provide context as to the interaction effects of seasonal, trend, and

background noise component, respectively. Models were constructed at not only the daily scale, but also three day, weekly, biweekly and monthly averages. Different temporal scales may return different importance and interaction effects between the algal biomass dynamics and parameters that may have a longer lag effect (*e.g.* T2m may not have a significant lag effect at the daily scale, but instead show greater importance at a longer temporal scale). The use of a DGBN can help conceptualize complex multivariate system interactions and identify key atmospheric or lake physical parameter contribution to algal biomass dynamics. The assessment of interaction effects may help better understand the dynamics of algal biomass in a changing climate.

6.2 Study Area

Daily to monthly mean lake physical and atmospheric forcing parameters are obtained for five North American Great Lakes; Great Bear Lake (GBL), Great Slave Lake (GSL), Lake Athabasca (LA), Lake Winnipeg (LW) and Lake Erie (LE), and their subsequent basins (Figure 6.1) over a total of 19 years (2002-2020). This research utilizes the same study area as outlined in section 4.2.
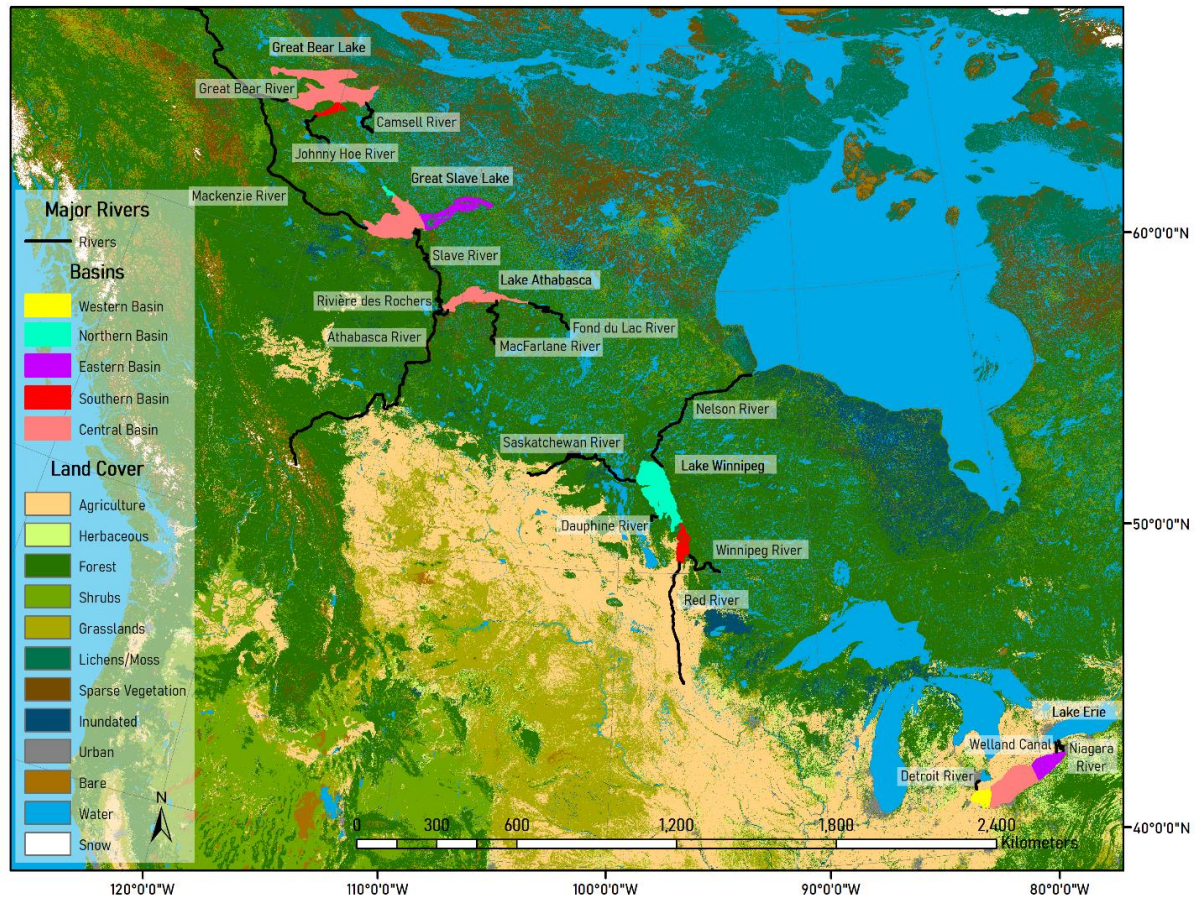
Figure 6.1. Map of study lakes and their corresponding basins and rivers. Land cover deliniation provided by ESA CCI Landcover (http://maps.elie.ucl.ac.be/CCI/viewer/download.php).

GBL is the largest of the selected study lakes, with a surface area of 31,153 km², a mean depth of ~76m (Johnson, 1975; Rouse *et al.*, 2008), and is comprised of two main basins; the larger, deeper Central basin (CB) (Smith, Dease, Keith, and McTavish arms), and the shallower south basin (SB) (McVicar arm) (Johnson, 1975). The CB of GBL is less productive with a mean annual chl-*a* concentration of 0.39 μg L^{-1}, from 2002-2020, while the SB remains more productive with a mean chl-*a* of 2.77 μg L^{-1} (section 4.5). The major inflows to GBL are the Camsell River (CB) and the Johnny Hoe River (SB), with the Great Bear River (CB) as the major outflow (Johnson, 1975; Chavarie *et al.*, 2015). GBL also exhibits the longest residence time of any of the selected study lakes at 124 years (Johnson, 1975), and is within a relatively undisturbed catchment (some mining operations and small settlements) (Schindler, 2001). The

136

mixing classification of GBL is that of cold monomictic; however, some regions are transitioning to a dimictic system (Woo & Rouse, 2007).

GSL is the second largest of the study lakes, with a surface area of 28,450 km^2 and is the deepest with a mean of ~88m (Rouse *et al.*, 2008). GSL is composed of three basins, the largest by surface area being the CB, the deepest waters of the east basin (EB), and the smallest and shallowest waters of the north basin (NB) (Avalon Rare Metals Inc, 2011; Shertzer *et al.*, 2008). The EB of GSL is the least productive of the three basins, with a mean annual chl-*a* concentration of 2.96 µg L^{-1} from 2002-2020, 4.49 µg L^{-1} in the CB, and 5.32 in the NB (section 4.5). The Slave River is the primary major inflow to GSL (CB), while the Mackenzie River is the primary major outflow (CB) (Evans, 2000; Gibson *et al.*, 2006). GSL exhibits the second longest residence time compared to the other study lakes at 16 years, however, is significantly lower than that of GBL (Evans, 2000; Gibson *et al.*, 2006). Similar to GBL, the catchment surrounding GSL is relatively undisturbed, with larger built-up regions (Yellowknife, Fort Providence, Fort Resolution). Yellowknife Bay of GSL has had an observed algal bloom in September 2013 (Pick, 2016) and exhibits a dimictic mixing cycle (Rouse *et al.*, 2008).

LA is the smallest of the selected study lakes with a surface area of 7,770 km^2, a mean depth of ~20m and has no discernable basins (Mitchell & Prepas, 1990). LA shows moderate primary productivity, with a mean chl-*a* concentration of 4.29 µg L^{-1} from 2002-2020 (section 4.5). There are many major inflow rivers that supply LA waters, such as the Athabasca River, Fond du Lac River, and MacFarlane River, with Rivière des Rochers as the primary outflow (Rasouli *et al.*, 2013). The surrounding catchment is largely undeveloped; however, there is oil sands wastewater deposited within the lake via the Athabasca River. Indigenous groups have attributed the observed increase in algal biomass within LA to oilsands operations (Bill *et al.*, 1994; Timoney, 2008; Parlee & D'Souza, 2019). Lake Athabasca has a residence time of ~6 years (calculated by dividing total LA volume by total inflow found in Bennett *et al.*, 1973), and exhibits a dimictic mixing cycle (Leconte *et al.*, 2008).

LW is the second smallest of the study lakes, with a surface area of 23,750 km^2, is the shallowest on average with a mean depth of ~12m and is comprised of two main basins: the

larger and deeper NB and the shallower and more productive SB (Government of Manitoba, 2021). The NB is slightly less productive, with a mean annual chl-*a* concentration of 7.38 μg L^{-1} from 2002-2020, while the SB is more productive with at 10.17 μg L^{-1} (section 4.5). Similar to LA, LW has many major river inflows, such as the Winnipeg River (SB), Saskatchewan River (NB), Red River (SB), and the Dauphine River (NB), with the Nelson River (NB) as the major outflow (Government of Manitoba, 2021; Warrack *et al.*, 2017). LW exhibits the second shortest residence time of ~3-5 years (Government of Manitoba, 2021). Unlikethe surrounding remote GBL, GSL, and LA, the catchment is significantly developed predominantly along the SB inflow rivers, with agriculture dominating land use, often leading to increase nutrient loading and algal blooms within the lake (Government of Manitoba, 2021; Binding *et al.*, 2018). Unlike the other lakes in the study, LW exhibits a polymictic mixing structure, where mixing events occur at irregular intervals (Nürnberg and LaZerte, 2016).

LE is the only Laurentian Great Lake selected for study due to its significant history with algal blooms (Ho & Michalak, 2015; Watson *et al.*, 2016; Smith *et al.*, 2015). LE has a surface area of 25,744 km2, and a mean depth of ~19m, with the shallowest occurring in the west end of the lake and the deepest regions in the Eastern portion (EPA, 2020). LE is comprised of three major basins, the shallowest and most productive west basin (WB), the moderately deep CB, and the deepest and EB. The WB is the most productive with a mean annual chl-*a* concentration of 4.04 μg L^{-1}, with the EB and CB being less productive at 2.83 μg L^{-1} and 1.27 μg L^{-1} respectively (section 4.5). The Detroit River is the main major inflow to LE, with the Niagara River and the Welland Canal as the main river outflows (Docker *et al.*, 2021). LE has the largest amount of development within the surrounding catchment of any of the studied lakes with a lengthy history of past algal bloom occurrences, the majority of which have occurred within the WB (Sayers *et al.*, 2019). LE EB and CB exhibit a dimictic mixing structure, while the shallower WB exhibits a polymictic mixing structure (Karatayev *et al.*, 2021).

6.3 Data Sources

6.3.1 ESA CCI Lakes

Complete daily lake-wide chl-*a* and LSWT values are derived from the ESA CCI Lakes plus dataset (v2.0.0 of the Lake Water Leaver Reflectance (LWLR) and the LSWT product), while other lake physical variables (LMLD, SNSR, SSR) and atmospheric forcings (T2m, WS, PPT, SR) are derived from the ECMWF ERA5-land data climate reanalysis product (2002-2020). The ESA CCI Lakes project is a harmonized lake data product that provides various water quality and quantity metrics (productivity, temperature, lake ice, water level, *etc.*), daily at a 1/120 latitude-longitude grid (ca.1 km)for more than 2,000 lakes globally. The ERA5-Land data product provides many climatic modelled reanalysis parameters (*i.e.*, T2m, LMLD, WS, *etc.*) hourly and a 9 km grid, globally.

The LWLR reflectance product, which provides daily estimated surface water chl-*a* concentrations, is developed using MERIS (2002-2012) and Sentinel-3A OLCI (2016-2020) satellite imagery. Within data version 2.0.0, there is currently no imagery between 2012-2016; current development looks to bridge the missing data using MODIS imagery. The chl-*a* estimates are calculated using atmospherically corrected images, with Idepix neural network routine to identify lake-wide extent and dynamic pixel classification (Carrea *et al.*, 2022). The atmospherically corrected images were matched with in situ observations within several days. The derived matched pairs were used to create several bio-optical algorithms for varying water types. Data may be missing due to cloud/ice cover and due to extreme solar zenith angles for high-latitude lakes during some winter months.

The LSWT was developed using a classification system to identify and exclude all non-water pixels via the visible-SWIR bands of the ASTR and AVHRR satellites and developed using a geophysical inversion of an optimal estimation technique. Multiple differing satellites are used to compare modelled and observed values, where inter-sensor adjustments are made using AVHRR as a reference. The LSWT product provides daily water skin temperatures

(<0.1mm thermal emission layer) at a spatial resolution of 1 km. Missing some daily data prior to 2006 is due to the lower coverage of ATSR at this time (Carrea *et al.*, 2022).

6.3.2 ERA5-Land Hourly Data

ERA5 is the fifth in a series of reanalysis products developed by ECMWF, originally consisting of atmospheric conditions in 1979 (FGGE) at a grid resolution of 208 km, the current iteration of ERA5 currently provides hourly atmospheric, land, and water parameter conditions from 1950-present at a 31 km grid resolution (Hersbach *et al.*, 2020). ERA5-Land provides improvements for terrestrial analytics by improving the resolution of ERA5 to a 9 km grid resolution (Muñoz-Sabater *et al.*, 2021). ERA5-Land utilizes the Tiled ECMWF Scheme for Surface Exchanges over Land (H-TESSEL) model, which incorporates ground- based (SYNOP) and satellite-derived weather, atmospheric, radiation, ocean and surface convection models.

The ERA5 2-m air temperature (T2m) represents the temperature (K) of the air as it is 2 m above the earth's surface. This product takes various atmospheric conditions into account and is calculated via an interpolation between the Earth's surface and the lowest H-TESSEL model level (Muñoz-Sabater *et al.*, 2021). The ERA5-Land data total precipitation (PPT, m) is modelled by the sum of both large-scale and convective precipitation patterns. The large-scale precipitation patterns are derived from the ECMWF integrated forecasting system (IFS) cloud scheme, where the formation of clouds is determined by changes in atmospheric conditions (pressure, temperature, and moisture). The convective precipitation is generated by the IFS convection scheme, which identifies convection at smaller scales. The modelled precipitation data do not factor fog, dew, or atmospherically evaporated water (Muñoz-Sabater *et al.*, 2021). The Surface Net Solar Radiation (SNSR, J m^{-2}) measures the amount of direct and diffuse shortwaves radiation (0.2-0.4 µm) at the earth's surface and subtracted by the surface albedo. Radiative flux is measured at the top-of-atmosphere (TOA) and considers both the absorption and scatter of radiance within the atmosphere due to clouds, aerosols, and atmospheric gases. Surface runoff (SR, m) is the water flux which exceeds infiltration rate, which is calculated using the TESSEL and H-TESSEL Models, taking into account the land use classification and

geometric relation soil class depth. Additionally, the modelled SR accounts for the vegetation and saturation points (Balsamo *et al.*, 2009; ECMWF, 2016). Similarly, subsurface runoff (SSR, m) utilizes both TESSEL and H-TESSEL models and accounts for subsurface water flux and soil water equation with varying soil layers (Balsamo *et al.*, 2009). Windspeed (WS, m s^{-1}) is measured in both Eastward (U) and Northward (V) 'neutral' wind speeds at 10 m above the Earth's surface, which accounts for surface roughness of inhomogeneous terrain. Neutral wind speeds assume that the air is stable and stratified, thereby providing slower speeds than real conditions during stable air masses and higher than real conditions during unstable air masses (ECMWF, 2016). The Lake Mixing Layer Depth (LMLD, m) indicates the depth at which freshwater has a uniform temperature gradient (Muñoz-Sabater *et al.*, 2021). The mixing layer depth is modelled through the use of the FLake model, which is done through the analysis of the lake temperature profile, wind speed, humidity, precipitation and both long and shortwave radiation (Betts *et al.*, 2020).

6.4 Methods

6.4.1 Preprocessing – ESA CCI Lakes

All ESA CCI Lakes data were downloaded locally from the CEDA archive. Satellite-derived chl-*a* approximation pixels may be missing or heavily influenced due to cloud cover, poor atmospheric conditions (*e.g.* haze), varying water optical types, satellite spatial-temporal coverage, and satellite positioning (*e.g.* high solar zenith angle). As such, various measures were taken to ensure the highest level of data integrity. Detailed descriptions of the initial preprocessing of the raw ESA CCI Lakes gridded data can be found in section 4.4, where per pixel outliers were detected temporally, ice covered and low confidence pixels masked, and interpolated temporally (linear). Spatial surface water chl-*a* and LSWT means within identified 1 km inward buffered lake basin boundaries were taken. Despite the removal of outliers and minimized variance from the spatial means, significantly higher algal biomass was observed during the ice on/off period, where few open water pixels were observed (Figure 6.2). Further preprocessing was done to the mean daily chl-*a* observations to minimize variance (Figure

6.2). Outlier detection of the daily mean time series (2002-2020) was calculated with the following equation:

$$O = Q3 + (1.5 * IQR)$$

where O is the outlier threshold, Q3 is the third quartile, and IQR is the interquartile range. Any daily value that exceeds this threshold was removed (converted to NA) from the time series. Due to the high ch-a approximations at the ice on/off period, daily observations where open water pixels were <90% of the maximum extent were removed (NA). Lake-specific growing seasons were identified in section 4.5.1 (Table 4.3), in which any chl-a and LSWT approximations outside of the growing season were converted to 0.00 μg L^{-1} and 273.15 K, respectively (the growing season was determined by average daily T2m from 2002-2020 where the winter season represents temperatures $\leq$ 273.15 K). Any missing observations (NA) were linearly approximated between the next available observations within the time series.
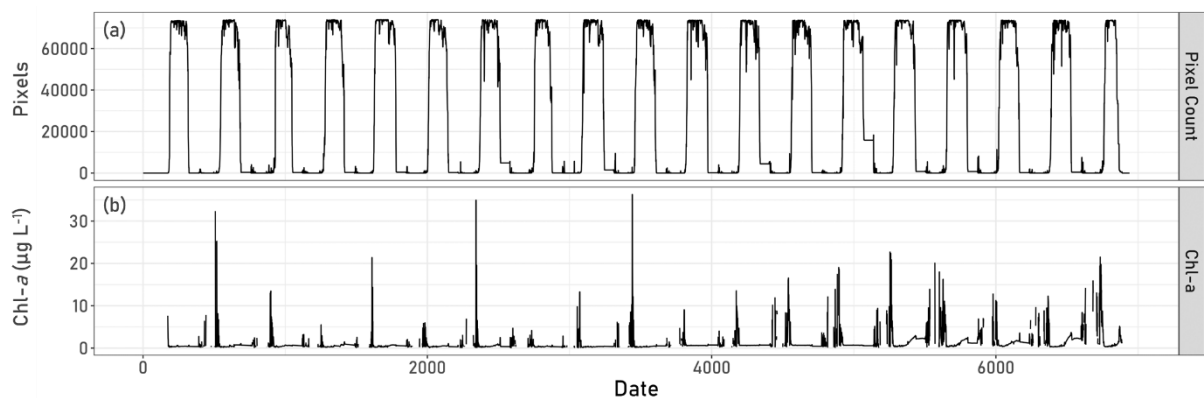
Figure 6.2. Line plots for GBL CB representing the (a) total daily count of open water pixels and (b) daily mean chl-a (μg L-1) concentrations after phase 1 preprocessing. Lower open water pixels are due to ice cover. Significant retrieval errors exist at the ice on/off period.

The CCI data version used in this study (2.0.0) has missing chl-a approximations from 2012-2015 due to the data gap between the MERIS satellite (2002-2012) decommission date, and its mission continuing satellite, Sentinel-3A (2016-present) launch date. New versions look to fill this gap, utilizing MODIS imagery. For this study, a complete temporal sequence is needed and therefore steps must be taken to fill this data gap. Daily chl-a from 2002 to 2011 was used to construct a multi-seasonal time series, where a univariate Holts-Winter method (base r 'stats' package) forecasted daily chl-a approximations from 2012 to 2015. All lakes

used the same alpha (α), beta (β) and gamma (γ) parameters (α = 0.05, β = 0.10, γ = NULL), with the exception of the Western basin of Lake Erie (α = 1.00, β = 0.10, γ = NULL). For the forecasted chl-*a* estimation (2012-2015) to match the rest of the time series, the same seasonal threshold was used to ensure 0.00 μg L^{-1} during the ice cover season. Refer to Figure 6.3 for a detailed list of steps.
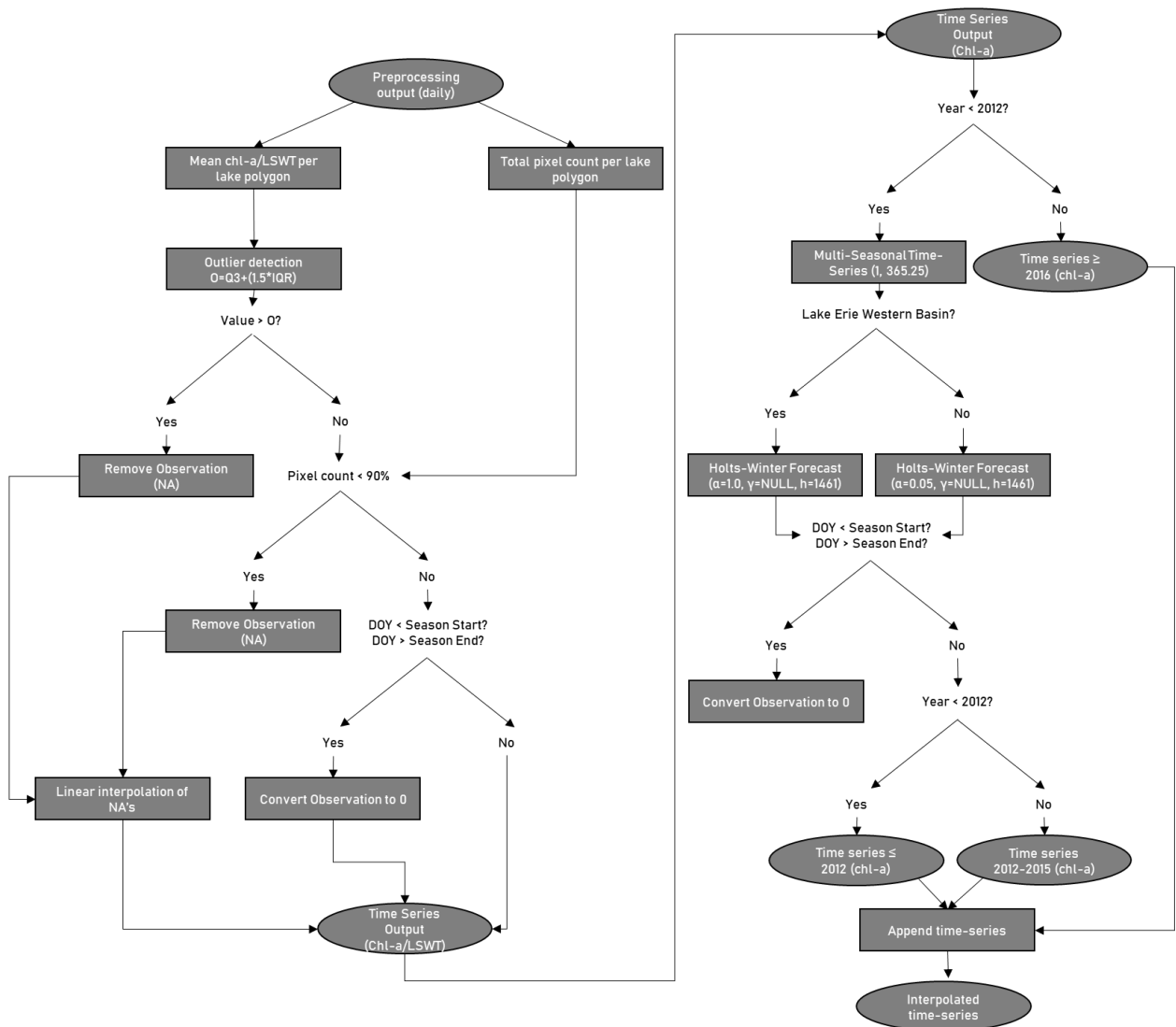
Figure 6.3. Daily time series processing of ESA CCI Lakes data. Daily raster pre-processing is as described in section 4.4.

6.4.2 Preprocessing – ERA5-Land

The ERA5-Land hourly data were downloaded from the Copernicus Climate Data Store (CDS) from 2002-2020 (Muñoz, 2019) utilizing a sub-region extraction (N: 69.05, E: -70.95, S: 38.95, W: -129.05). The hourly gridded data was used to calculate a daily mean T2m, LMLD, SNSR and WS, while daily sums were calculated for PPT, SR and SSR. Spatial means within 1km inward buffered lake basin boundaries were taken daily for T2m, LMLD, SNSR and WS; however, PPT, SR, and SSR were derived from the surrounding catchment (Figure 6.4). Since the ERA5 data are a forward model reanalysis product, sensor failure and approximation variance are not likely to impact the time series integrity and therefore no additional preprocessing was needed.

6.4.3 Watershed Flow Pathways

The impact of PPT, SR, and SSR on algal biomass dynamics is not typically due to what is occurring within the lake boundaries but instead within the surrounding catchment, due to the mobilization of organic and inorganic matter. Daily PPT, SR, and SSR are calculated as a spatial mean within the watershed boundary. Since the lakes are divided into different basins, the watershed must be delineated to determine which areas flow into which basins. The original watershed extent is defined by using the level-04 HydroSHEDS product (Lehner *et al.,* 2008). Flow accumulation and direction was calculated using a 1 km digital elevation model (CEC, 2007) in ESRI ArcGIS. The fill tool (spatial analyst) was used to fill local minimum elevations, while the flow direction tool (spatial analyst) is used to determine the steepest slope direction in a D8 pathway. The flow accumulation tool (spatial analyst) calculates the total number of prior gridded cells that flow toward the target. The flow accumulation raster was converted to polylines by first reclassifying to a binary raster where 0 = no flow and 1 = flow (accumulation thresholds vary depending on the size of the watershed). The stream link tool (spatial analyst) utilizes the reclassified raster and the direction raster to calculate the flow direction and stream orders. The stream-to-feature tool (spatial analyst) is used to convert the rasters to a polyline, where the line symbology can provide an arrow indicating directionality as seen in Figure 6.4. From here the watershed was manually delineated where regions which show direct

intersection with a given lake basin were associated with said basin. These regions define where the PPT, SR and SSR spatial means are derived.
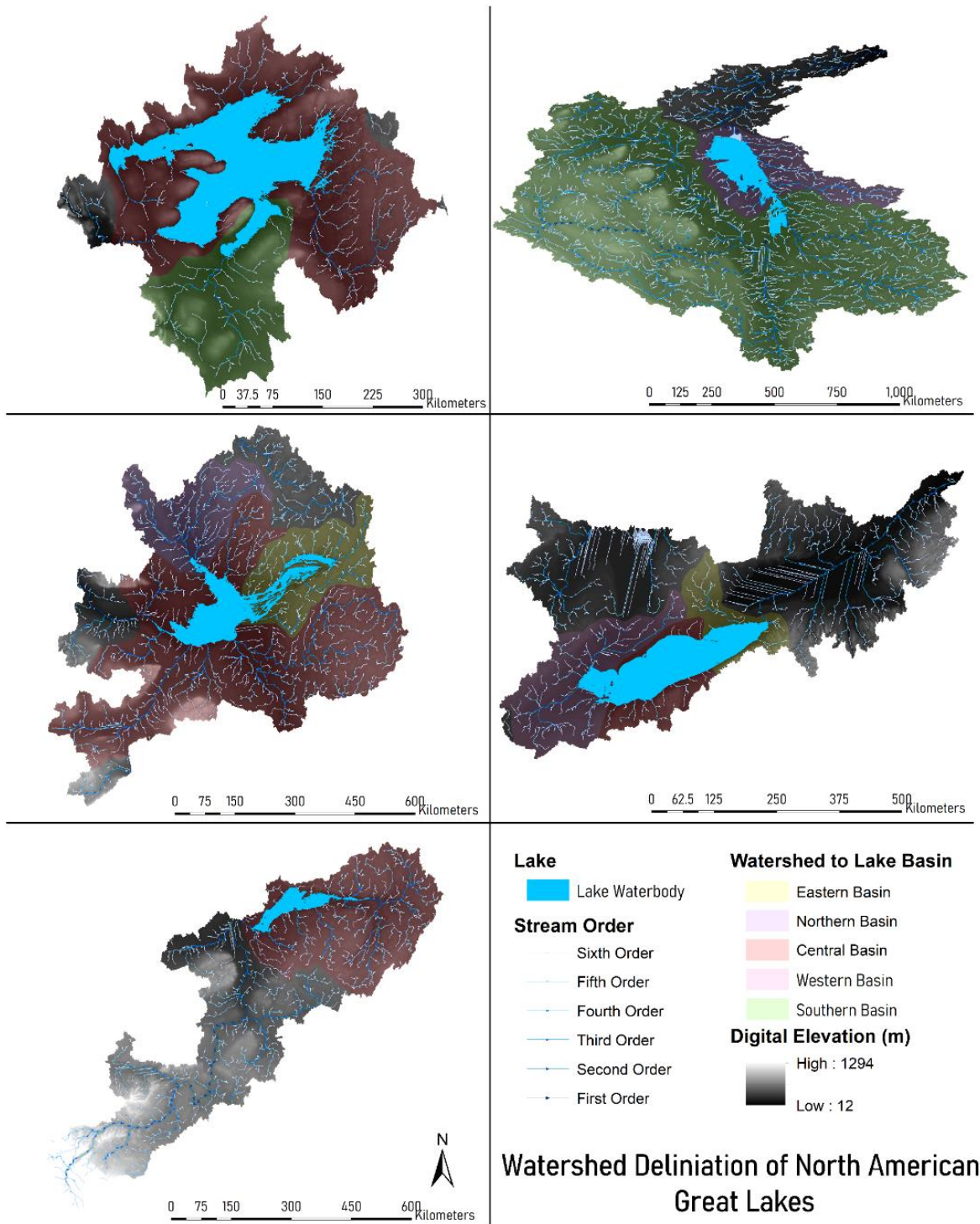
Figure 6.4. Watershed deliniation defined by the flow direction and accumulation of water associated intersecting with a lake basin. DEM provided by CEC, 2007, and watershed boundaries provided by hydroSHEDS (Lehner *et al.,* 2008).

146

6.4.4 High-Order Dynamic Gaussian Bayesian Network

Bayesian Networks (BN) are a statistical analysis method used to infer conditional dependencies amongst variables (Nagarajan *et al.,* 2013). Traditionally, BNs are trained using discrete data and are comprised of two components: a directed acyclic graph (DAG) and a probability distribution (Sebastiani, *et al.,* 2009). The DAG provides a visual representation of the network, comprised of nodes and arcs (Nagarajan *et al.,* 2013). A node (or variable) that is conditionally independent is considered a parent node and is represented by a marginal probability table. A node that is conditionally dependent is considered a child node and, therefore, represented by a joint probability table. The arcs are directed moving from a parent node to a child node, which is conditionally dependent on the parent. The structure of the network can be determined using priors (requiring user expertise) or learned through various optimization methods (*i.e*., Hill-Climbing, Particle Swarm Optimization (PSO), K2, etc.).

Adaptations have been made for BNs with continuous data known as a Gaussian Bayesian Network (GBN), where the parent nodes represent a Gaussian distribution, and child nodes a conditional multivariate Gaussian distribution (Nagarajan et. al. 2013; Neapolitan, 2004). As this study utilizes only continuous data, GBN form the foundation of parent-child relationship. The GBN assumes normality of the data in the global network; however, some studies have relaxed this assumption using non-parametric BNs (Vitale *et al.,* 2020; Marella *et al.,* 2019). The parent-child node relationship is represented as a linear Gaussian model, as defined by Quesada *et al*. (2021a):

$$p(x_j|Pa_j) = \mathcal{N}(\beta_{0j} + \beta_{1j}x_{1(j)} + \cdots + \beta_{rj}x_{r(j)}; \sigma_j^2) \tag{6.1}$$

Where x_j represents the target node, $Pa_j = \{x_{1(j)}, \dots, x_{r(j)}\}$ is the set of parent nodes for x_j, $\mathcal{N}$ represents the density function of a normal distribution, β_{0j} is the independent coefficient (intercept), $\{\beta_{1j}, \dots, \beta_{rj}\}$ are the parent coefficients (slope), while σ_j^2 is the variance (error) of x_j. The parent-child node connections are represented by multivariate Gaussian linear regressions; however, this method differs from frequentist approaches (*e.g*., OLS). The GBN fits distributions based on a fixed sample with random parameters compared to the fixed

coefficients of OLS (Law *et al.,* 2014), and are therefore considered more flexible with high variance data (Ickstadt *et al.,* 2011; Rhodes, 2017; StataCorp, 2023). The GBN can be used to estimate a target node based on the network coefficients derived from conditional dependencies using a Monte Carlo simulation and are therefore more applicable to high dimensional networks where the multivariate distribution is complex (Nagarajan *et al.,* 2013; Koller & Friedman, 2009). Provided that all nodes are linear Gaussian models, the joint Gaussian distribution of the network can be written as described by Quesda *et al.* (2021a):

$$p(x) = \prod_{i=1}^{n} p(x_j | Pa_j) = \prod_{i=1}^{n} \mathcal{N}\left(\beta_{0j} + \sum_{j=1}^{r(i)} \beta_{ji}\, x_{j(i)}; \sigma_j^2\right) \tag{6.2}$$

Where $x = (x_1, \dots, x_n)$ represents the number of nodes within the network, and $r(i)$ represents the total amount of parent nodes for x_j. The model coefficients are estimated using a multivariate linear regression Gaussian distribution. Once the network is constructed, the network can be used to inference new data based on the distributions of the parent nodes. The BN and GBN methods are, however, static, and do not consider the temporal domain. Dynamic GBN (DGBN) is utilized to determine the interaction effects of variables as they change over time (Nagarajan *et al.*, 2013). The structure of a DGBN is similar to that of the GBN, however, each variable is described in the temporal domain $X = \{X_i(t); i = 1, \dots, k; t = 1, \dots, t\}$. Variables are also conditionally dependent on the past observations of other variables by a given time-step (Nagarajan *et al.*, 2013). The joint probability of the DGBN now includes all time slices t to a given horizon T as described by Quesada *et al.* (2021a):

$$p(X_0, X_1, \dots, X_T) = P(X_{0:T}) = p(X_0) \prod_{t=0}^{T-1} p(X_{t+1} | X_{0:t}) \tag{6.3}$$

Where $X_t = (X_{1t}, X_{2t}, \dots, X_{nt})$ is the vector of all the nodes in a time slice t, where $t = 0, 1, \dots, T$ and T is the maximum number of time slices. It is often assumed that the temporal process is homogeneous and invariant, which is often not true with most real-world data (such as those provided in this study), and instead used to simplify the model (Nagarajan *et al.,* 2013). Flexible models that relax this assumption typically require a larger number of observations

148

(Xuan Vinh *et al.,* 2012). Studies have successfully implemented DBGN with non-homogenous data (Quesada *et al.,* 2021a; Grzegorczyk, 2015). Most implementations of DGBNs are first-order Markovian, in which the state of a given time slice is dependent on only the preceding time slice (Koller and Friedman, 2009). Surface water algal biomass, however, may have a longer lagged response to an environmental stressor than one just on day, such as a strong mixing event (*i.e.*, precipitation), which may initially disrupt algal growth but then increase the biomass via nutrient mobilization (Reynolds, 1994). Given the potential for longer lag effects, high order Markovian models (known as a high order DGBN, or HO-DGBN) are needed. HO-DGBN exponentially increase the complexity of the network and therefore require significant training time (Quesada *et al.,* 2021a). To reduce model complexity, this study implemented transition networks, where the arcs are directed to only the most recent time slice of chl-*a* (the target of interest) (Quesada *et al.,* 2021b). A HO-DGBN can be represented as explained by Pasquini Santos *et al.* (2014), where X_t in equation_ is $X_t = \{X_t^1, X_t^2, ..., X_t^N\}$ for N variables in T total time slices, and the conditional independencies for the probability distribution is $p(X_{t+1}, X_{t+2}, ..., X_{t+T}|X_t)$. From here the network can be restricted so that directed arcs are only calculated from the present time slice to any other time slice (inter-time slice), disregarding arcs within a time slice.

This research uses an HO-DGBN, implementing a transition network to reduce training time using the R package 'dbnR' by Quesada and Valverde (2022), which also uses functions from the R package 'bnlearn' (Scutari *et al.,* 2023). The HO-DGBN was optimized using a Particle Swarm Optimization (PSO) method with order invariant encoding, which directed interactions to only the most recent time slice, improving training time (Quesada *et al.,* 2021b). Data from 2002-2019 were used to train the model, where data from 2020 were reserved for testing. All training data was normalized by the following equation:

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{6.4}$$

and the forecasted values were reconverted by adding the mean (μ) and standard deviation (σ) of the original distribution to the forecasted chl-*a*. To determine the optimal number of time-

slices, the HO-DGBN was trained from 1-5 days. Each iteration forecasted the 2020 data using an approximation Markov-Chain Monte Carlo simulation 10 times. Each iteration calculated the Root Mean Squared Error (RMSE), where the lowest RMSE was recorded. The network structure with the lowest RMSE was selected for analysis. Increasing the number of time-slices exponentially increases training time, and so a max of 5 days was implemented. However, a lag effect may persist at a longer interval than what is captured within 5 days. To increase the temporal range of the lag search without increasing training time, block averages are taken for all the input data every three days, weekly, biweekly and monthly. The change in temporal frequency (*i.e.*, decreasing the temporal resolution by moving from daily to monthly frequencies) is represented by Δ. These temporal frequencies were tested for the optimal lag order in the same method. The assumption of homogeneity may also affect forecast results, as the provided data is non-homogeneous. To determine the impacts of the non-stationary signal, the timeseries were decomposed into their stationary, non-stationary and residual signals, where an HO-DGBN was constructed for each, along with the complete time series. Performance metrics between temporal signal time series indicated not only which signal introduced the higher error, but provided information on the dynamics of the season, trend, and the residuals. The decomposed time series were reconstructed by taking the sum of the same forecasted chl-*a* concentration from each temporal signal, as seen in the following equation:

$$\widehat{Chla}_{reconstructed} = \widehat{Chla}_{stationary} + \widehat{Chla}_{nonstationary} + \widehat{Chla}_{residuals} \qquad (6.5)$$

The reconstructed chl-*a* concentrations provided insight into whether time series decomposition can improve chl-*a* forecasts compared to those without decomposition (complete time series). A total of 220 HO-DGBNs were constructed, in which each recorded the RMSE, normalized RMSE (NRMSE), and Mean Absolute Error (MAE) (Table 6.1).

Table 6.1. Summary of performance and error metrics for the HO-DGBN

Parameter	Equation	Variables	Measure	EQ (#)
RMSE	$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}$	n = sample size, y_i = the observed value, $\hat{y}_i$ = the predicted value	Standard deviation of the prediction errors	6.6

Parameter	Equation	Variables	Measure	EQ (#)
NRMSE	$NRMSE = \dfrac{RMSE}{\sigma}$	$RMSE$ = Root Mean Square Error, σ = standard deviation of input dependent data	RMSE normalized by the standard deviation of the input training data. Allows for comparison between models of varying values	6.7
MAE	$MAE = \dfrac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n}$	n = sample size, y_i = the observed value, $\hat{y}_i$ = the predicted value	Average error between pairs of predicted and observed values	6.8
MAD	$MAD = \dfrac{1}{n}\sum_{i=1}^{n}|y_i - \bar{y}|$	n = sample size, y_i = the observed, $\bar{y}$ = mean of the dataset	Determines the absolute deviation of an observed value from the true mean	6.9

The network structure provided insight into which parameters contribute to the algal biomass, while the parent-child node coefficients indicate the nature of the interaction (*i.e.*, positive vs negative) at a given timestep. To better understand the coefficients across the many temporal frequencies, signals, and time-steps, the average positive and negative coefficients were calculated across all lakes. The absolute sum of all parent node coefficients for each parameter was divided by the absolute sum of all parent nodes coefficient in the network to determine each parameters percentage of contribution made to present time-slice chl-*a* ($chl - a_0$). The greater the contribution of a given parameter, the greater its significance in the HO-DGBN.

To determine the importance of a given variable within the network, an out-of-bag (OOB) approach was implemented. Using the selected model for each lake basin, temporal frequency and signal, a different input variable (apart from chl-*a*), was iteratively excluded from the network. Each HO-DGBN model was forecasted 10 times, where the median MAE was recorded. Models where the MAE is highest given a parameter is excluded indicated it was the most important variable, while the lowest MAE indicated it was least important. The sensitivity of the model was tested by iteratively changing one variable in the testing data by a given percentage (-30%, -15%, -5%, 0%, 5%, 15%, 30%) of the normalized value, and forecasting the new chl-*a* concentration. To determine which variable in the model was most sensitive to change, the Mean Absolute Deviation (MAD) was calculated for each percent change, where the forecast with no change (0%) is used to calculate the mean of which the other forecasted values may deviate from. The sensitivity not only provides insight into which

input parameters are more sensitive to change, but also what the algal response may be given that change, which could infer our understanding of how algal dynamics might be affected under future climate scenarios.

6.5 Results

6.5.1 Time-Series

Of the nine parameters (Figure 6.5), LSWT, Chl-*a* and SR most frequently returned a Theil-Sen slope of 0 across all lakes (Figure 6.6), due to the inclusion of the winter season, where chl-*a* is an assumed 0.00 µg L^{-1} and LSWT is 273.15 K. Across the change in temporal frequency (Δ), most parameters did not return a significant ($p < 0.05$) Theil-Sen slope (Table 6.2). The number of significant Theil-Sen slopes typically decreased with the increase in Δ, however its magnitude tended to increase, most notably with LSWT, PPT, SR, and chl-*a* in LE (Figure 6.6). SSR most frequently returned significant Theil-Sen slopes (45.45% – 81.82% of lakes), where all lakes and basins show a general decrease (-2.68^{-6} m yr^{-1} – 5.80^{-7} m yr^{-1}) apart from LE WB, LE CB and the GBL CB (1.10^{-7} m yr^{-1} – 2.41^{-6} m yr^{-1}). The mean rate of change in T2m for Northern Lakes (NL; GBL, GSL, LA) (0.08 K yr^{-1}) was twice that of the Southern Lakes (SL; LW, LE) (0.04 K yr^{-1}) across all temporal frequencies. NL saw an average decrease in LMLD (-0.01 m yr^{-1}) compared to the increase seen in SL (0.01 m yr^{-1}) across all temporal frequencies.

152
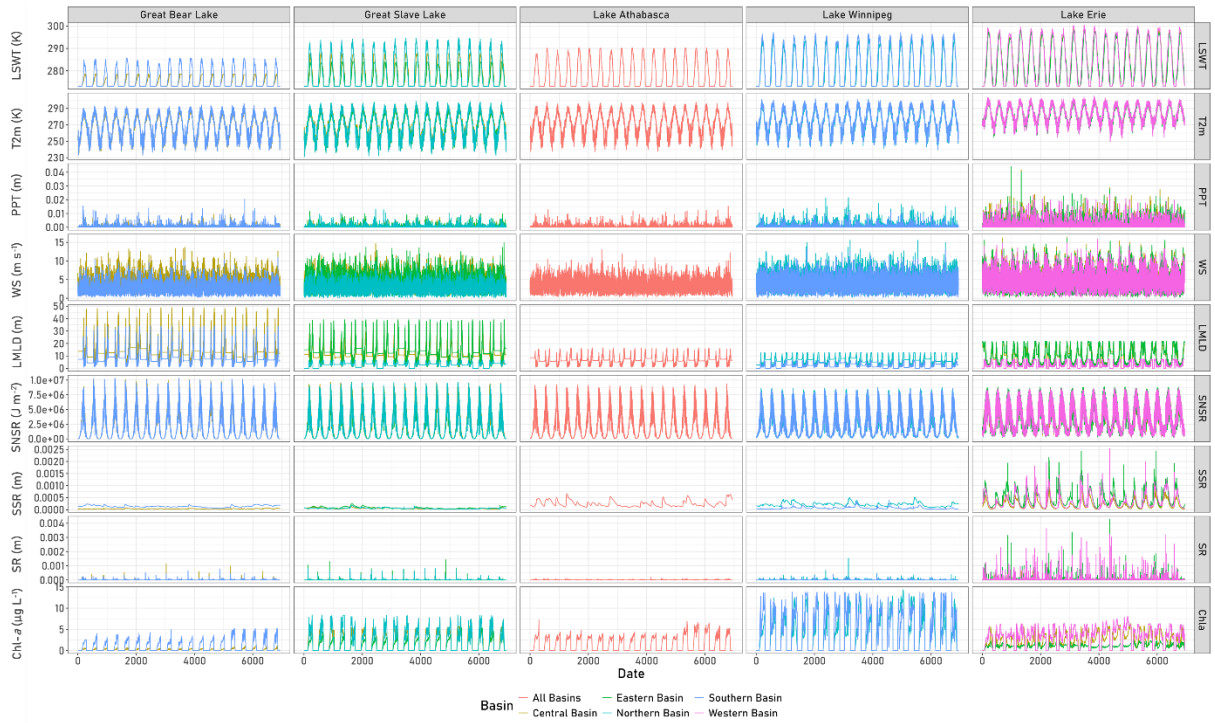
Figure 6.5. Daily time series from January 1, 2002, to December 31, 2020, for Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR), and Chlorophyll-a (Chl-a).
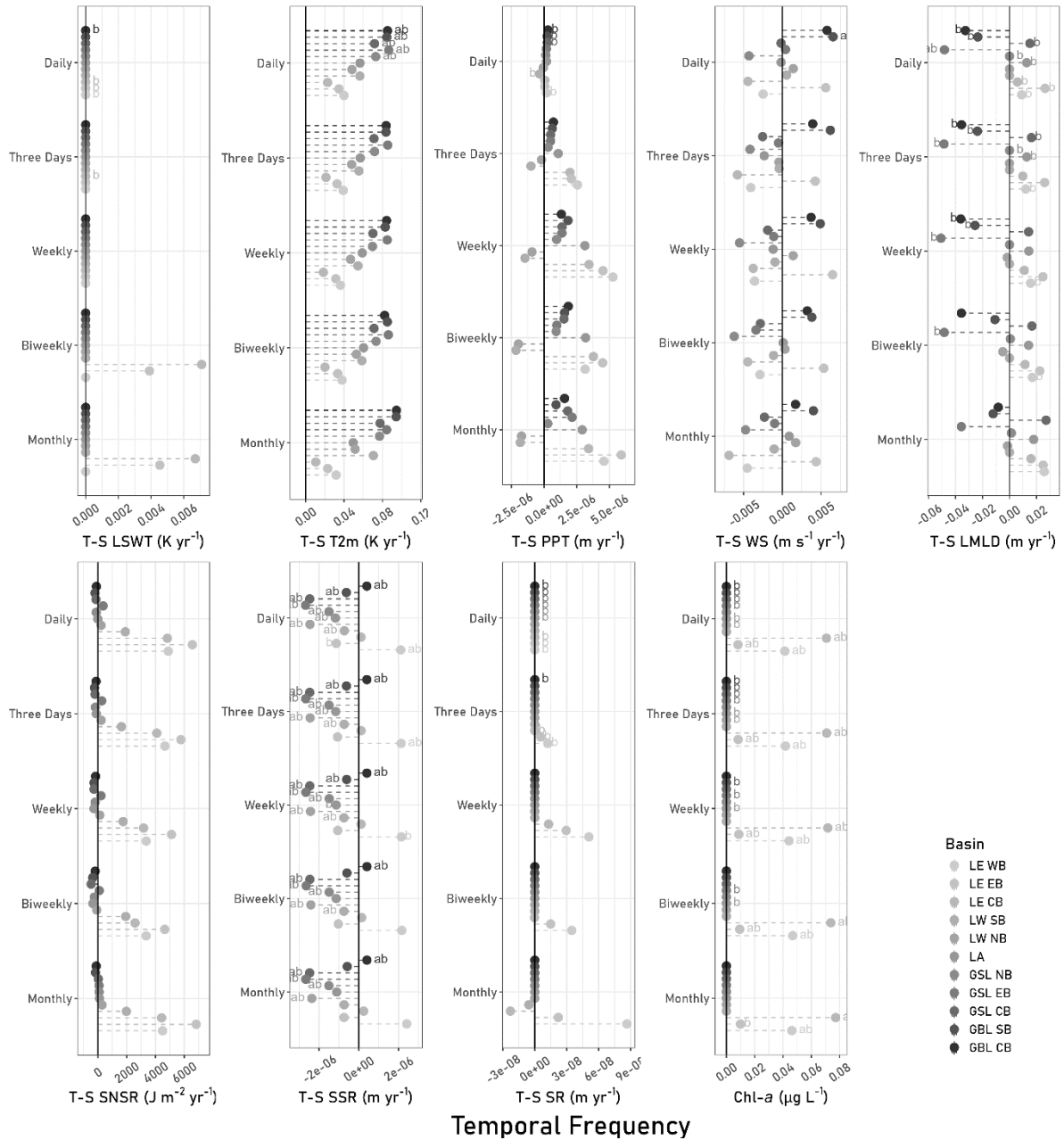
Figure 6.6. Theil-Sen (T-S) slope for each aggregated temporal frequency and lake basin. Calculated T-S slopes were converted to an annual metric to create a common unit. The letter 'a' represents a T-S p-value of <0.05, and the letter 'b' represents a Mann-Kendall p-value of < 0.05. Daily n = 6940, 3-day n = 2314, weekly n = 992, biweekly n = 496, monthly n = 228. LSWT = Lake Surface Water Temperature, T2m = 2m Air Temperature, PPT = Total Precipitation, WS = wind speed, LMLD = Lake Mixing Level Depth, SNSR = Surface Net Solar Radiation, SSR = Subsurface runoff and Chl-a = Chlorophyll-a.

154

Table 6.2. Percentage of lakes (n = 11) that exhibit a significant (p < 0.05) Theil-Sen slope for each HO-DBNR input parameter, from daily to monthly temporal block means. LSWT = Lake Surface Water Temperature, T2m = 2m Air Temperature, PPT = Total Precipitation, WS = wind speed, LMLD = Lake Mixing Level Depth, SNSR = Surface Net Solar Radiation, SSR = Subsurface runoff and Chl-a = Chlorophyll-a.

Time	LSWT	T2m	TP	WS	LMLD	SNSR	SSR	SR	Chl-*a*
Daily	0.00%	45.45%	0.00%	9.09%	9.09%	0.00%	81.82%	0.00%	27.27%
Three Days	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	81.82%	0.00%	27.27%
Weekly	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	63.64%	0.00%	27.27%
Biweekly	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	54.55%	0.00%	27.27%
Monthly	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	45.45%	0.00%	18.18%

Of the nine parameters for use in the HO-DGBN, SNSR was the only parameter where no lakes exhibited a significant Mann-Kendall trend for any Δ, as SNSR is unlikely to exhibit a significant change over a 19-year period. The Mann-Kendall Test had found significant trends for the majority of lakes with the LMLD (0.00% – 81.82%), SSR (45.45% – 90.91%), SR (0.00% – 81.82%) and Chl-*a* (27.27% – 81.82%) parameters (Table 6.3). As Δ decreased, so did the number of lakes with significant Mann-Kendall trends. The lack of significant trends indicates a minimal influence of the non-stationary signal in the HO-DGBN.

Table 6.3. Percentage of lakes (n = 11) that exhibit a significant (p < 0.05) Mann-Kendall trend for each HO-DBNR input parameter, from daily to monthly temporal block means. LSWT = Lake Surface Water Temperature, T2m = 2m Air Temperature, PPT = Total Precipitation, WS = wind speed, LMLD = Lake Mixing Level Depth, SNSR = Surface Net Solar Radiation, SSR = Subsurface runoff and Chl-a = Chlorophyll-a.

Time	LSWT	T2m	TP	WS	LMLD	SNSR	SSR	SR	Chla
Daily	36.36%	45.45%	45.45%	9.09%	81.82%	0.00%	90.91%	81.82%	81.82%
Three Days	9.09%	0.00%	0.00%	0.00%	63.64%	0.00%	81.82%	36.36%	81.82%
Weekly	0.00%	0.00%	0.00%	0.00%	36.36%	0.00%	81.82%	0.00%	54.55%
Biweekly	0.00%	0.00%	0.00%	0.00%	18.18%	0.00%	54.55%	0.00%	45.45%
Monthly	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	45.45%	0.00%	27.27%

6.5.2 Lag Order

Of the 55 HO-DGBN constructed using a stationary signal (11 lake basins, 5 temporal frequencies), the majority of lakes returned lower NRMSE when having a higher number of time slices, with 85.5% (47/55) returning an optimal T of 4-5, with only 14.5% (8/55) returning

an optimal T of 1-3. For the non-stationary, 29.1% (16/55) returned an optimal T of 4-5, and 70.9% (39/55) returned an optimal T of 1-3. For the residuals, 47.2% (26/55) returned an optimal T of 4-5, and 52.7% (29/55) returned an optimal T of 1-3. Without decomposition, 65.5% (36/55) returned an optimal T of 4-5, and 34.5% (19/55) returned an optimal T of 1-3 (Figure 6.7). The T in Figure 6.7 are the those selected for each HO-DGBN moving forward.
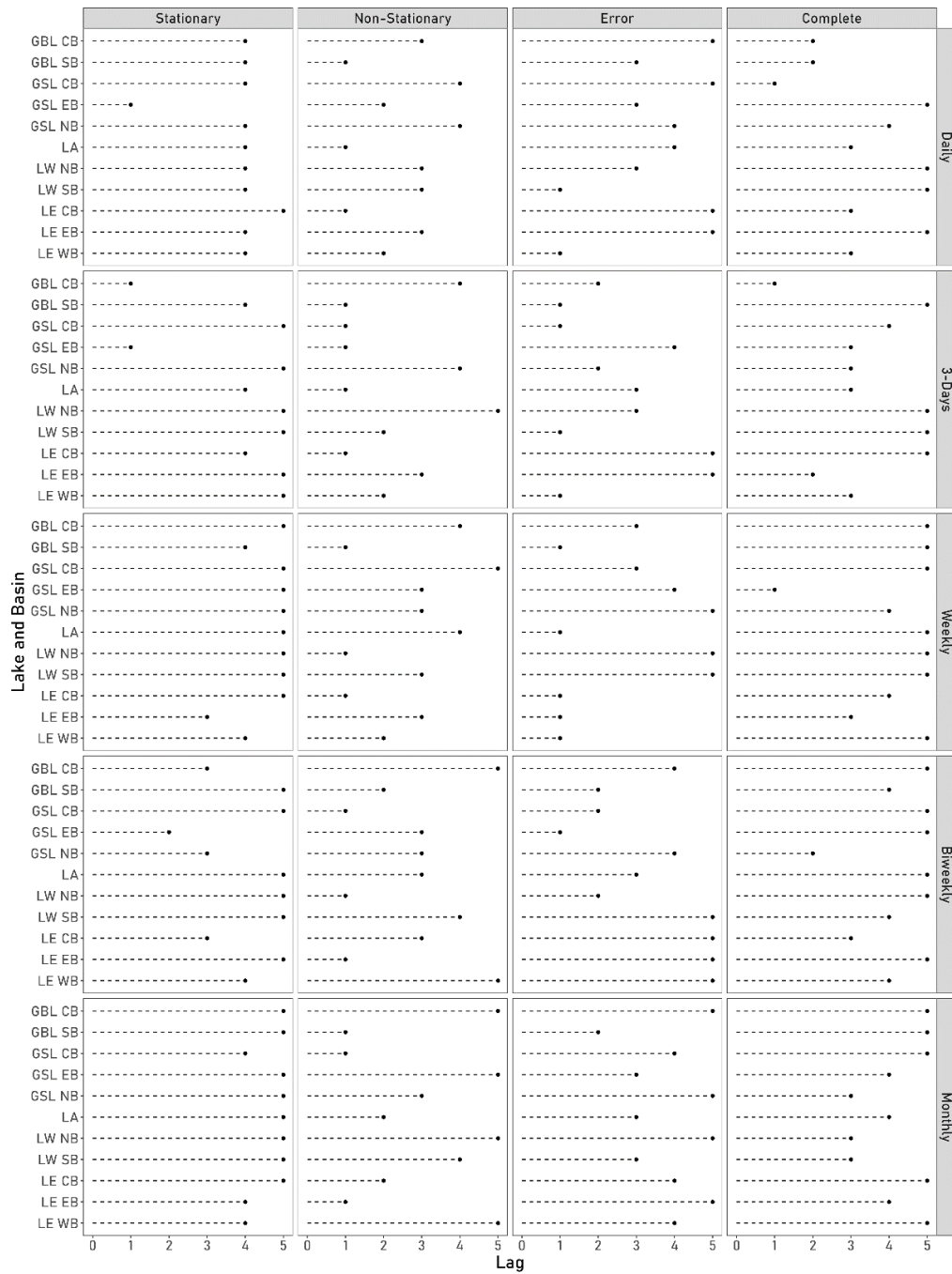
Figure 6.7. HO-DGBN models across 1-5 time-steps with the lowest forecasting error (NRMSE) for 5 North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin and WB = West Basin. HO-DGBN models are constructed using the original time series (complete), the stationary temporal signal, the non-stationary, and the time-series residuals. Each HO-DGBN is also constructed using a daily time series, along with a three-day block mean, weekly, biweekly and monthly means.

157

6.5.3 Performance Between Temporal Signals and Frequencies

The HO-DGBN using only the stationary signal returned the lowest error ($\widehat{NRMSE}$ = 0.31) compared to the non-stationary the residuals, complete, and the reconstructed ($\widehat{NRMSE}$ = 1.09, 0.93, 0.65, and 0.42 respectively) signals, across all temporal frequencies (Figure 6.8). The Δ returned a statistically significant (p < 0.1, n = 5) linear decrease in error for the stationary (β = -0.06 – -0.13 NRMSE Δ^{-1}) and complete (β = -0.05 – -0.13 NRMSE Δ^{-1}) signals for all lakes. Despite the poor performance of the non-stationary and residual only models, reconstructing the forecasted chl-*a* concentrations from a decomposed time series returned a lower error compared to no decomposition for daily to biweekly ($\widehat{NRMSE}$ difference = 0.001 – 0.43) frequencies, while the monthly ($\widehat{NRMSE}$ difference = -0.15 – 0.20) frequency returned higher error for 4 lake basins (GSL CB, GSL EB, LW NB and LE EB).
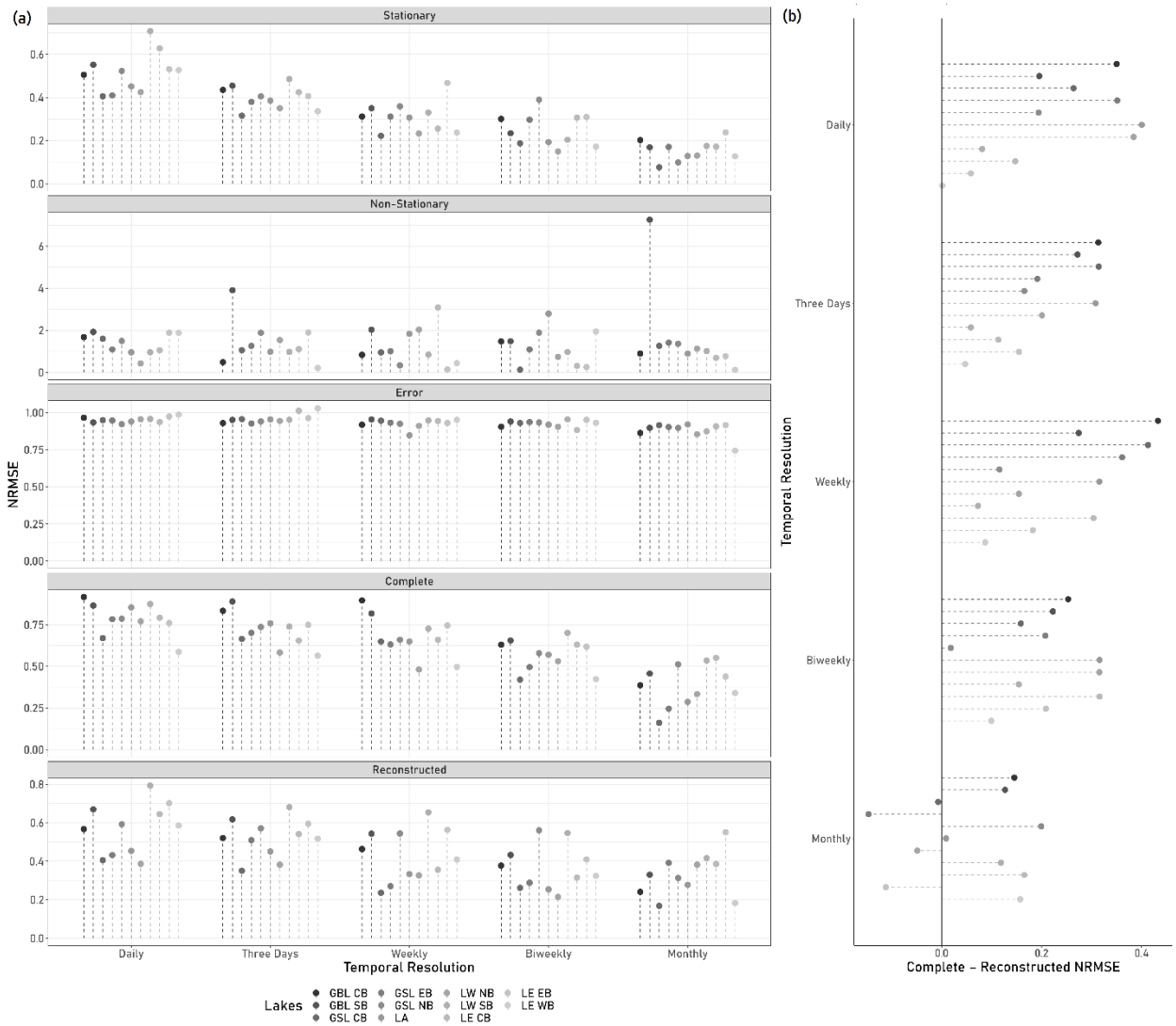
Figure 6.8. (a) Performance results of the HO-DGBN models with the lowest time-step forecasting error (NRMSE). (b) Difference between the NRMSE of the forecasted chl-a constructed with the original time-series (complete) HO-DGBN and the NRMSE of the reconstructed forecasted chl-a from the stationary, non-stationary and residual HO-DGBN. Models are created for five North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin, and WB = West Basin.

6.5.4 Performance Between Lakes

The HO-DGBN using the stationary signal across all Δ returned a similar forecasting error between NL and SL ($\widetilde{NRMSE}$ = 0.31 and 0.31, respectively). When using the non-stationary signal, the SL returned lower error compared to the NL ($\widetilde{NRMSE}$ = 0.97 and 1.07

respectively), primarily driven by the poor forecasting results for GBL SB ($\widetilde{NRMSE}$ = 2.03). Notably, LE non-stationary chl-*a* forecasts at the monthly Δ compared to all other lakes ($\widetilde{NRMSE}$ = 0.69 and 1.20 respectively). The HO-DGBN models constructed with time series residuals, the original time series, and reconstructed time series returned no significant differences between NL and SL ($\widetilde{NRMSE}$ = 0.43 – 0.93 and 0.41 – 0.95, respectively).

There was little difference in model performance between lake regions. Models constructed from only the stationary signals were the best performing; however, both complete and reconstructed models were also adequate in forecasting chl-*a* concentrations, primarily at the monthly temporal scales (Figures 6.9 and 6.10). Therefore, HO-DGBN models constructed with daily data provide networks with poorer structure, and therefore the results pertaining to the algal biomass dynamics are less accurate compared to the monthly model data. The results provided indicate that the HO-DGBN method can provide an accurate forecasting of algal biomass as they change over time, with improved accuracy with an increase in Δ. These models can, therefore, be used to understand which parameters were the most important in understanding algal biomass, the nature of the interactions, and the model sensitivity to a change in the parameters.
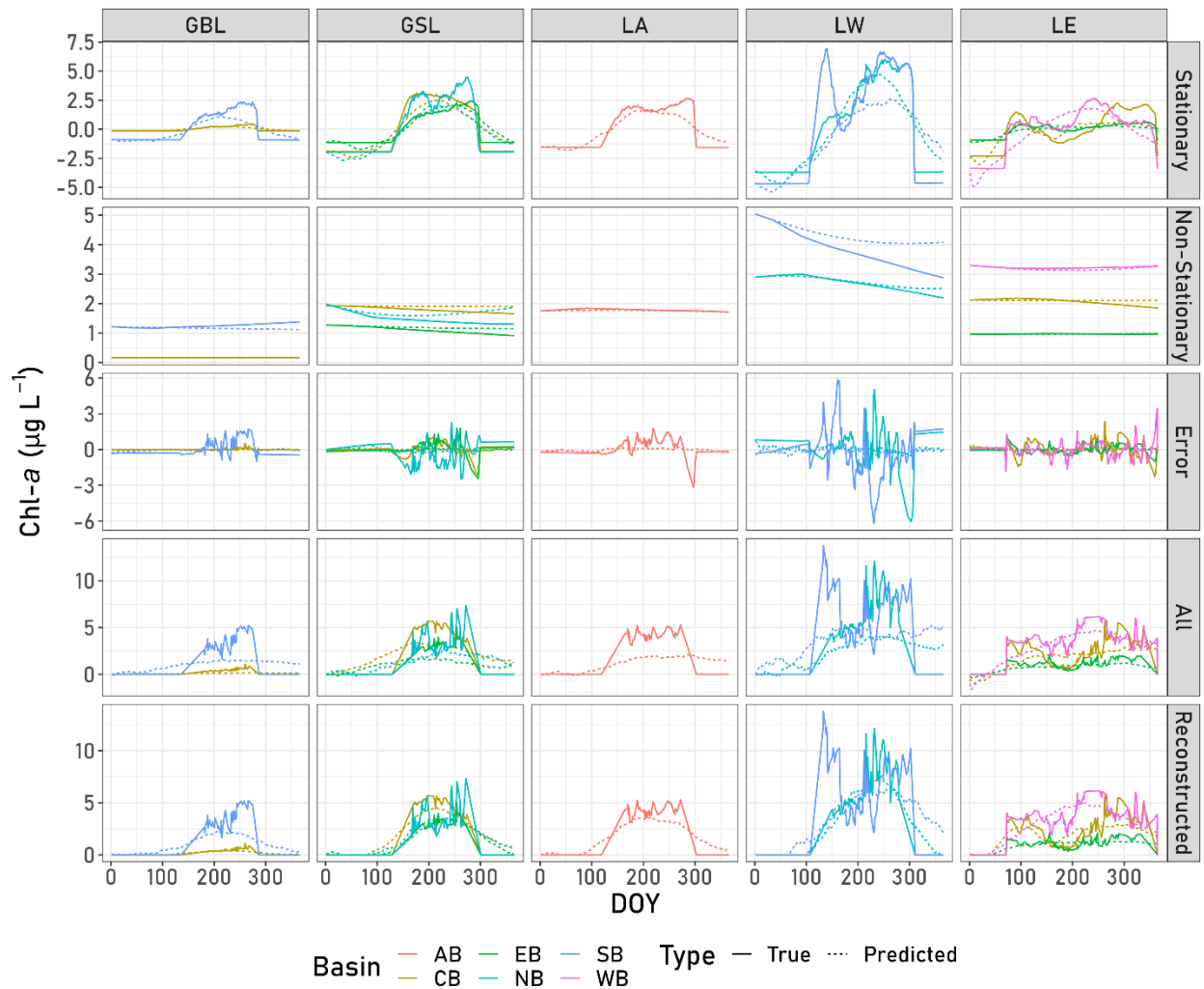
Figure 6.9. Forecasted chl-a concentrations derived from the HO-DGBN using daily time series inputs for 5 North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin and WB = West Basin. HO-DGBN models are constructed using the original time series (all), the stationary temporal signal, the non-stationary, and the time-series residuals.
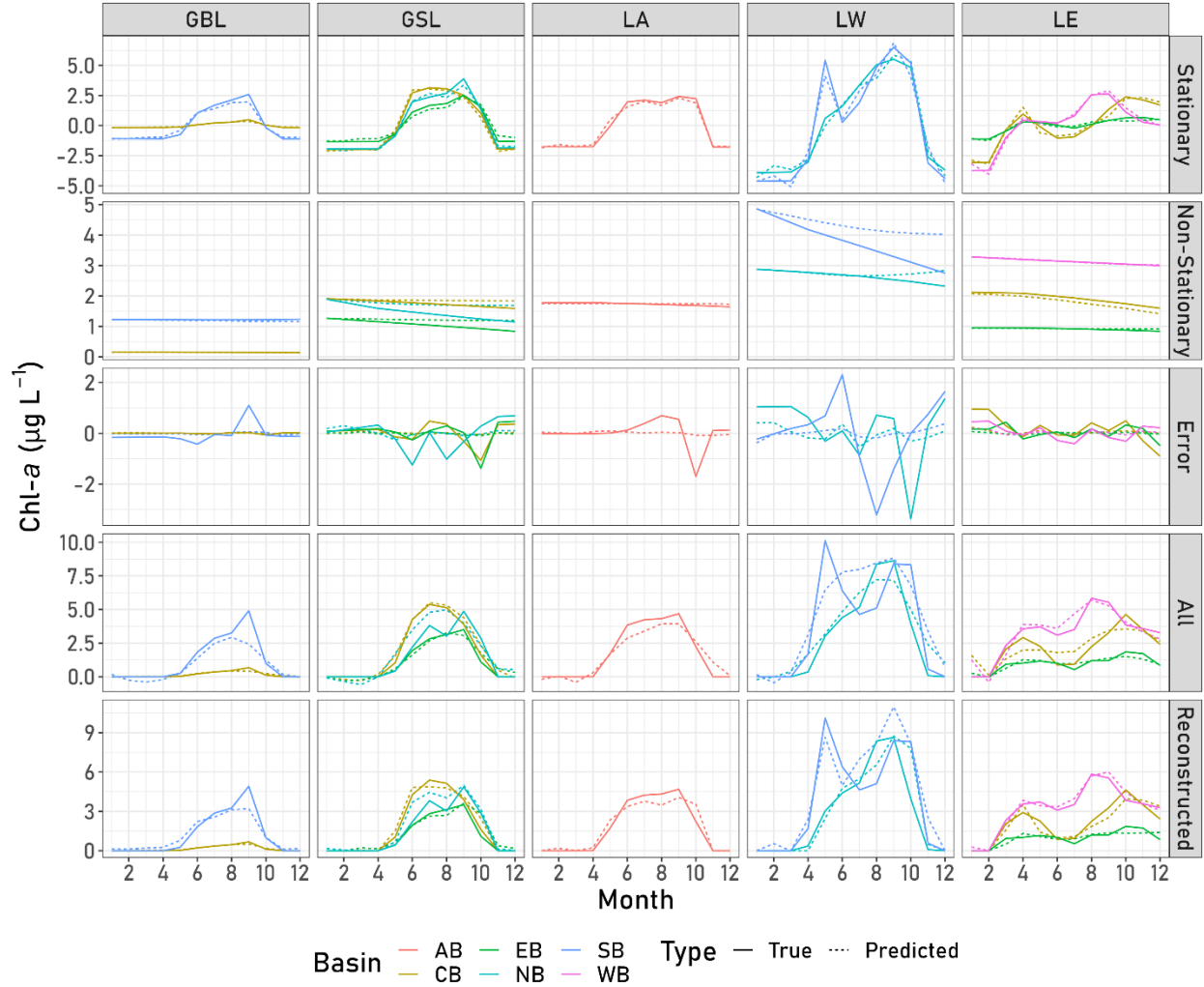
Figure 6.10. Forecasted chl-a concentrations derived from the HO-DGBN using monthly block mean time series inputs for five North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin, and WB = West Basin. HO-DGBN models are constructed using the original time series (all), the stationary temporal signal, the non-stationary, and the time-series residuals.

6.5.5 Variable Importance

Across all lakes and temporal frequencies, LSWT returned the highest forecasting error ($\widehat{MAE} = 6.19\ \mu g\ L^{-1}$) when excluded from the stationary HO-DGBN, followed by T2m, SNSR, SSR, LMLD, SR, WS, and PPT the lowest ($\widehat{MAE} = 1.01\ \mu g\ L^{-1}$) (table 6.4). The non-stationary HO-DGBN returned highly variable results with no clear pattern across all lakes and temporal frequencies, with LMLD returning the highest error ($\widehat{MAE} = 0.18\ \mu g\ L^{-1}$) followed by LSWT,

T2m, SSR, PPT, WS, SNSR and SR the lowest ($\widehat{MAE}$ = 0.11 μg L^{-1}). Similarly, the HO-DGBN constructed using the residuals of the time series returned minimal variance in results between all lakes and temporal frequencies, with SSR returning the highest error when excluded ($\widehat{MAE}$ = 0.50 μg L^{-1}), followed by SNSR, SR, WS, LMLD, T2m, LSWT and PPT the lowest ($\widehat{MAE}$ = 0.43 μg L^{-1}). Without any decomposition, LSWT returned the highest error ($\widehat{MAE}$ = 1.70 μg L^{-1}) when excluded from the HO-DGBN model followed by T2m, SNSR, LMLD, SSR, SR, PPT and WS the lowest ($\widehat{MAE}$ = 1.10 μg L^{-1}). When the forecasted chl-*a* were reconstructed, LSWT returned the highest forecasting error ($\widehat{MAE}$ = 6.25 μg L^{-1}) when excluded from the HO-DBGN models (stationary, non-stationary and residual), followed by SNSR, T2m, SSR, LMLD, SR, PPT and WS the lowest ($\widehat{MAE}$ = 1.25 μg L^{-1}).

Table 6.4. Median MAE of variable importance across all temporal frequencies and lakes. Variable importance uses an out-of-bag approach to determine which variable when excluded from the HO-DGBN returns the highest error. All values represent a forecasted chl-a concentration MAE (μg L^{-1}).

	Stationary	**Non-Stationary**	**Error**	**Complete**	**Reconstructed**
LSWT	6.19	0.16	0.44	1.70	6.25
T2m	3.48	0.16	0.45	1.60	2.82
PPT	1.01	0.14	0.43	1.11	1.37
WS	1.26	0.13	0.46	1.10	1.25
LMLD	1.64	0.18	0.45	1.18	1.48
SNSR	2.99	0.12	0.47	1.41	3.08
SSR	2.42	0.16	0.50	1.16	2.27
SR	1.39	0.11	0.46	1.11	1.37

With the increase in Δ across all lakes for the stationary, complete, and reconstructed HO-DGBN, LSWT remains the most important parameter, while all other parameters show a shift in importance (Figure 6.11). Stationary HO-DGBN models show increasing importance for SR, SSR and PPT as Δ increases, while SNSR, T2m, LMLD and WS decrease in importance, with the reconstructed HO-DGBN models returning similar results (Figure 6.12). For the non-stationary HO-DGBN models, PPT, SNSR, SSR, and SR returned an increase in importance with increasing Δ, while LSWT, T2m, WS, and LMLD decreased in importance. Residual HO-DGBN returned decreasing importance for all parameters.
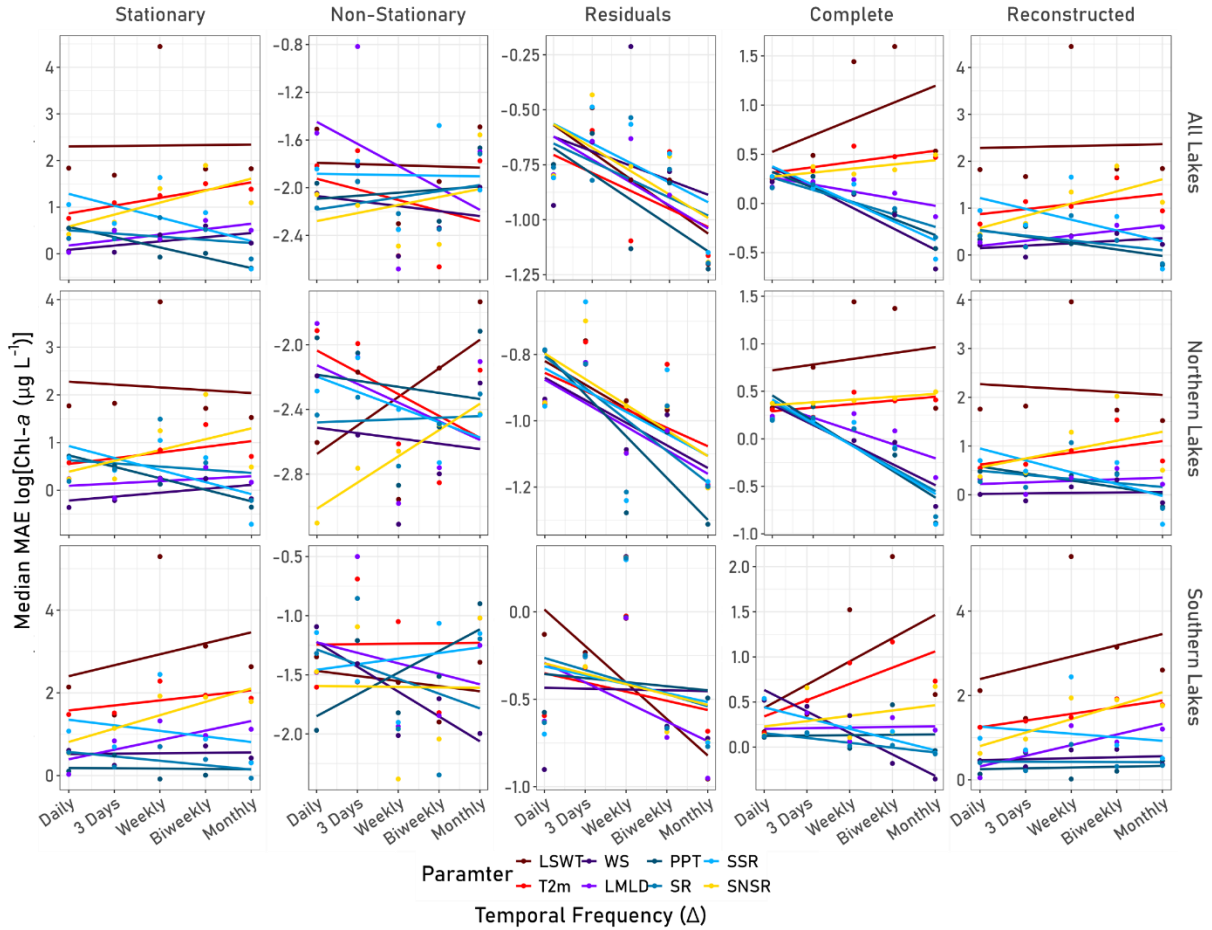
Figure 6.11. Median MAE of forecasted chl-a derived from HO-DGBN importance sampling across all study lakes as the change with increasing temporal frequency (from HO-DGBN construced using daily data to those using monthly block means). Importance sampling utilizes a mutilated network to exclude a given parameter. Higher forecasting error of chl-a when excluded indicates greater importance to the network. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR).

164
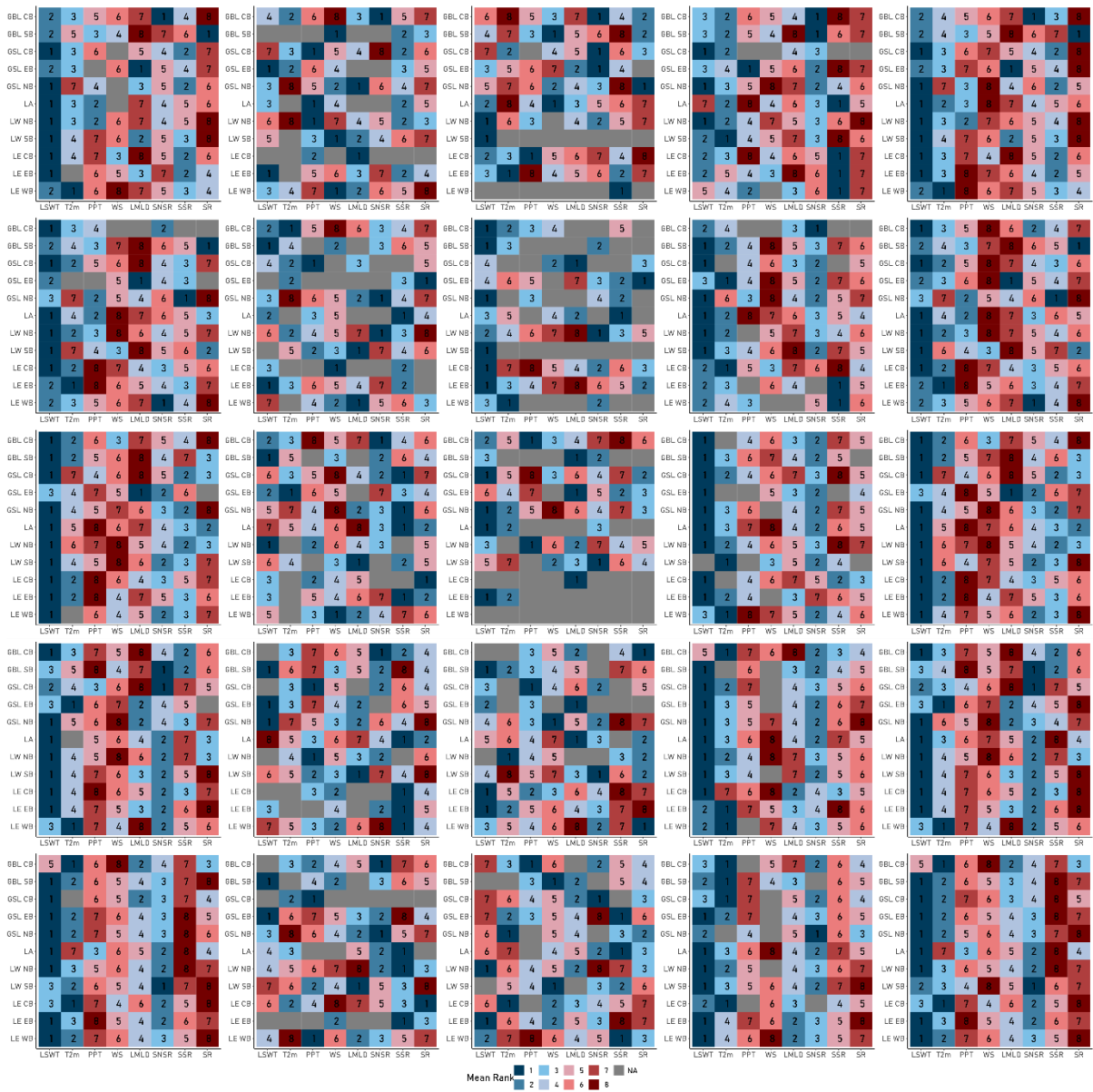
Figure 6.12. Ranked parameter importance determined using a mutilated network importance sampling. Ranks are determined by the MAE derived from forecasting chl-a when a given parameter is excluded per lake. A rank of 1 indicates the highest importance (highest MAE), while a rank of 8 indicates the lowest importance (lowest MAE). NA indicates that no parent nodes of the parameter were used in the HO-DGBN network; therefore, it cannot be excluded for importance sampling. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR). HO-DGBN models are constructed for five North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg, and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin and WB = West Basin. HO-DGBN models are constructed using the original time series (complete), the stationary temporal signal, the non-stationary, the time-

165

series residuals, and forecasted chl-a reconstucted from the decomposed time series (reconstructed). Each HO-DGBN is also constructed using a daily time series, along with a three-day block mean, weekly, biweekly and monthly means.

6.5.6 Model Coefficients

The HO-DGBN provides a transparent network where the modelling coefficients can be extracted to understand how each parent node is weighted in its relation to $chl - a_0$. A breakdown of the mean modelling coefficients across all lakes per temporal signal and Δ can be found in Figure 6.13.
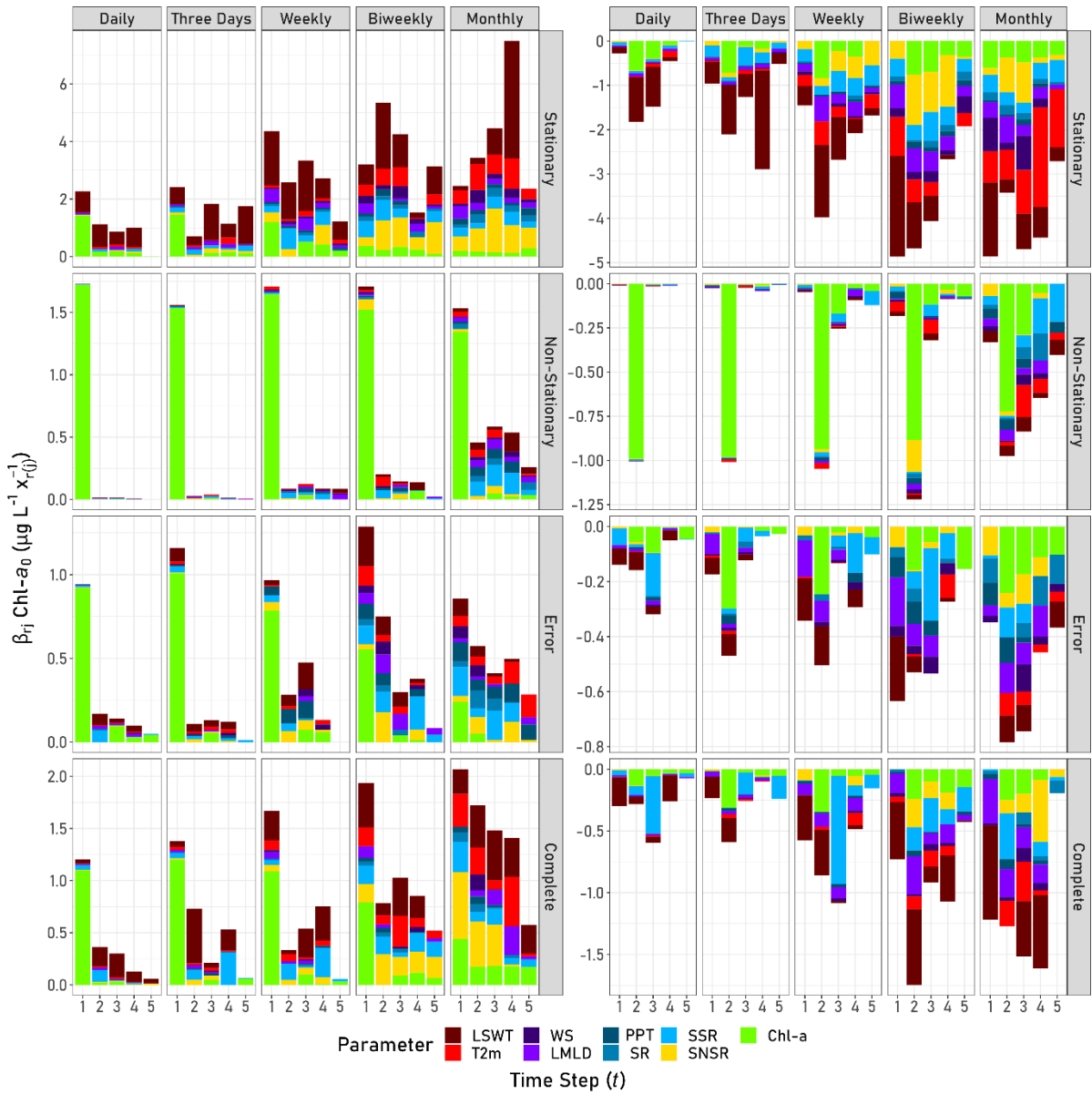
Figure 6.13. Average HO-DGBN model coeffecients (β) of the parent nodes per time-step (rj) for child node $chl - a_0$ across all study lakes. HO-DGBN models are constructed using the original time series (complete), the stationary temporal signal, the non-stationary, and the time-series residuals. Each HO-DGBN is also constructed using a daily time series, along with a three-day block mean, weekly, biweekly and monthly means. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR) and Chlorophyll-a (Chl-a).

167

6.5.6.1 Stationary

Across all lake basins, time-steps and Δ for the stationary HO-DGBN, LSWT was the largest weighted parameter on average, contributing to 38.32% of the normalized β_{rj} for $chl - a_0$, with past observations of chl-a at 16.49%, SNSR at 11.42%, T2m at 9.87%, SSR at 9.70%, LMLD at 6.83%, WS at 3.41%, SR at 2.24% and PPT as the lowest, contributing 1.72%. As Δ increased (daily to monthly), both LSWT and past observations of chl-a decreased in β_{rj} contribution, while SNSR, SR, T2m, PPT and WS all increased with Δ. Both LMLD and SSR increased with Δ, which peaked at the weekly scale and then decreased with increasing Δ (Figure 6.14). The relationship between the input parameters and $chl - a_0$ were mixed (both positive and negative depending on T). Past observations of chl-a were primarily positive at the daily to weekly Δ, while primarily negative from the biweekly to monthly Δ. LMLD and WS were also primarily negative from the biweekly to monthly Δ. Across all parameters and lake basins from the daily to monthly Δ, the majority of β_{rj} contribution occurred within a t of 1-2. For daily to weekly Δ, the majority of the positive β_{rj} contribution occurred at $t = 1$, while biweekly was at $t = 2$, and monthly was at $t = 4$. For daily – weekly Δ, the majority of the negative β_{rj} contribution occurred at $t = $ 2-4, while biweekly and monthly was at $t = 1$ (Figure 6.13).

6.5.6.2 Non-Stationary

Across all lakes basins, time-steps and Δ for the non-stationary HO-DGBN, past observations of chl-a were overwhelmingly the largest weighted parameter on average, contributing to 74.04% of the normalized β_{rj} for $chl - a_0$, with SSR at 6.02%, LSWT at 3.66%, T2m at 3.60%, LMLD at 3.11%, SNSR at 3.08%, PPT at 2.80%, SR at 2.30%, and WS as the lowest, contributing 1.40%. As Δ increased (daily to monthly), only past observations of chl-a showed a decrease in β_{rj} contribution while LSWT, T2m, LMLD, WS, SR, SSR, PPT and SNSR all increased with Δ. Similar to the stationary signal, most parameters exhibit a mixed positive to negative β with chl-a contributing a mostly positive β at the daily to weekly Δ, and mostly negative from biweekly to monthly Δ. Across all parameters and lake basins

from the daily to monthly Δ, the majority of the positive β_{rj} contribution occurred at $t = 1$, while the majority of negative β_{rj} contribution occurred at $t = 2$.

6.5.6.3 Residuals

Across all lakes basins, time-steps and Δ for the residual HO-DGBN, past observations of chl-a were the largest weighted parameter on average, contributing to 38.98% of the normalized β_{rj} for $chl - a_0$, with LSWT at 14.92%, SSR at 12.68%, LMLD at 7.78%, PPT at 6.62%, SNSR at 6.40%, T2m at 4.87%, SR at 3.98% and WS as the lowest, contributing 3.77%. As Δ increased (daily to monthly), both past observations of chl-a and LSWT decreased in β_{rj} contribution, while T2m, WS, LMLD, PPT, SR, and SNSR all increased in β_{rj} contribution with Δ. SSR, however, returned the lowest contribution at the three-day Δ and peaked at the biweekly Δ (figure 6.13). Each parameter in the residual HO-DGBN model returned varying positive and negative effects on $chl - a_0$. The past observations of chl-a, SNSR, and T2m returned a primarily positive β, LMLD, SR and WS were primarily negative, while LSWT, SSR and PPT returned a mixed effect. Across all parameters and lake basins from the daily to monthly Δ, the majority of the positive β_{rj} contribution occurred at $t = 1$. The majority of negative β_{rj} contribution varied by Δ, where three day, weekly and monthly peaked at $t = 2$, while daily peak at $t = 3$, and biweekly peaked at $t = 1$ (Figure 6.13).

6.5.6.4 Complete

Across all lakes basins, time-steps and Δ for the complete HO-DGBN, LSWT was the largest weighted parameter on average, contributing to 28.05% of the normalized β_{rj} for $chl - a_0$, with past observations of chl-a at 27.17%, SSR at 17.93%, SNSR at 8.86%, LMLD at 6.99%, T2m at 6.77%, PPT at 1.56%, SR at 1.34% and WS as the lowest, contributing 1.32%. As Δ increased (daily to monthly), past observations of chl-a, LSWT and SSR returned a decrease in β_{rj} contribution, while T2m, LMLD, WS, SR, PPT and SNSR increased in β_{rj} contribution with Δ. Each parameter in the complete HO-DGBN returned varying interactions with $chl - a_0$, where past observations of chl-a, SNSR and T2m returned primarily positive

169

β_{rj} contribution, while LMLD returned a primarily negative β_{rj} contribution. Parameters LSWT, SR, SSR, PPT and WS returned mixed β_{rj} contribution depending on the temporal lag and lake. Across all parameters and lake basins from the daily – monthly Δ, the majority of the positive β_{rj} contribution occurred at $t = 1$. The majority of negative β_{rj} contribution varied with Δ where daily and weekly Δ peaked at $t = 3$, three-day and biweekly Δ peaked at $t = 2$, while monthly Δ peaked at $t = 4$. All β_{rj} contribution percentages can be seen in Table 6.5 and Figure 6.14.

Table 6.5. Proportion of summed coeffecient β across all lakes and basins that exhibit either a positive or negative value at a given time slice (**T**), for each temporal signal (stationary, non-stationary, residuals (error) and the complete time series) at each temporal frequency (daily to monthly). Colours represent the cell value (greens = high propotion, reds = low proportion).

		Positive					Negative				
		$T=1$	$T=2$	$T=3$	$T=4$	$T=5$	$T=1$	$T=2$	$T=3$	$T=4$	$T=5$
Stationary	Daily	24.33%	12.09%	9.36%	10.72%	0.17%	2.96%	19.58%	15.85%	4.84%	0.10%
	Three Days	15.45%	4.56%	11.74%	7.34%	11.27%	6.15%	13.51%	8.08%	18.57%	3.33%
	Weekly	16.71%	9.91%	12.80%	10.41%	4.68%	5.57%	15.22%	10.26%	7.98%	6.45%
	Biweekly	8.98%	15.00%	11.92%	4.31%	8.78%	13.64%	13.13%	11.38%	7.47%	5.39%
	Monthly	6.09%	8.50%	11.05%	18.57%	5.86%	12.06%	8.49%	11.66%	11.00%	6.73%
Non-Stationary	Daily	61.41%	0.59%	0.55%	0.22%	0.00%	0.38%	35.78%	0.61%	0.44%	0.00%
	Three Days	56.72%	1.04%	1.40%	0.48%	0.26%	0.94%	36.61%	0.82%	1.49%	0.24%
	Weekly	46.78%	2.34%	3.36%	2.40%	2.30%	1.33%	28.69%	6.93%	2.56%	3.31%
	Biweekly	41.55%	4.89%	3.53%	3.36%	0.55%	4.46%	29.68%	7.77%	2.09%	2.11%
	Monthly	23.37%	6.95%	8.90%	8.17%	3.96%	5.03%	14.86%	12.76%	9.86%	6.14%
Error	Daily	44.62%	8.01%	6.67%	4.68%	2.34%	6.55%	7.49%	15.04%	2.36%	2.23%
	Three Days	49.09%	4.62%	5.52%	5.15%	0.51%	7.35%	19.91%	5.21%	1.48%	1.16%
	Weekly	29.95%	8.71%	14.71%	4.11%	0.00%	10.59%	15.60%	4.12%	9.07%	3.13%
	Biweekly	26.14%	15.24%	6.07%	7.68%	1.72%	12.89%	10.74%	10.84%	5.54%	3.14%
	Monthly	16.08%	10.80%	7.73%	9.34%	5.35%	6.53%	14.72%	13.97%	8.58%	6.90%
Complete	Daily	33.75%	10.15%	8.42%	3.58%	1.65%	8.38%	7.93%	16.76%	7.27%	2.11%
	Three Days	31.77%	16.82%	4.86%	12.25%	1.56%	5.40%	13.60%	5.91%	2.32%	5.50%
	Weekly	25.62%	5.14%	8.28%	11.58%	0.85%	8.86%	13.21%	16.62%	7.45%	2.37%
	Biweekly	19.35%	7.83%	10.26%	8.53%	5.19%	7.27%	17.43%	9.16%	10.70%	4.28%
	Monthly	15.83%	13.18%	11.34%	10.79%	4.39%	9.31%	9.74%	11.60%	12.34%	1.48%
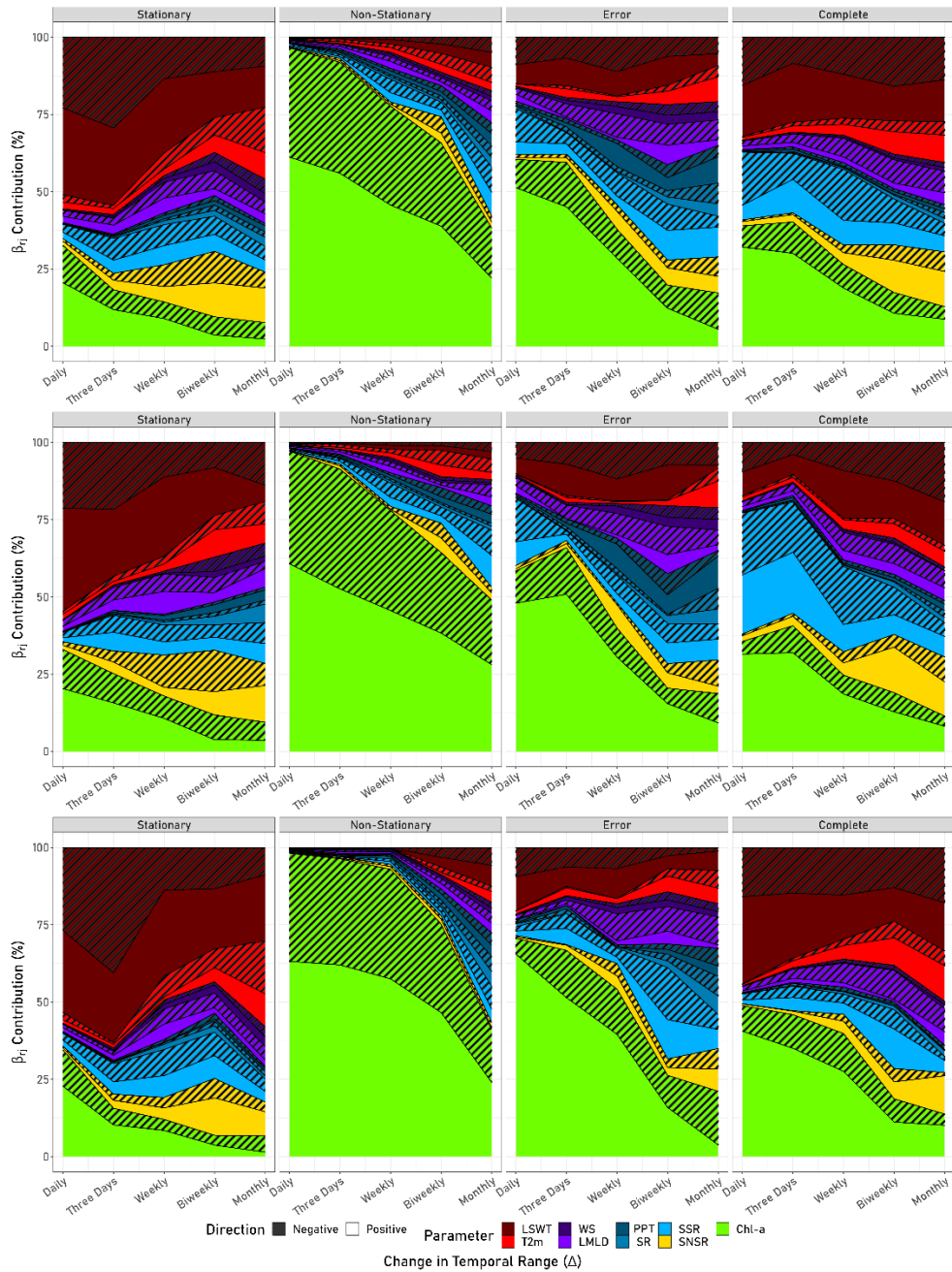
Figure 6.14. Average HO-DGBN model coeffecients ($\boldsymbol{\beta}$) of the parent nodes per time-step ($\boldsymbol{rj}$) for child node $\boldsymbol{chl-a_0}$ across all study lakes and time-steps, represented as a percentage of the sum of the absolute $\boldsymbol{\beta}$. Shaded regions indicate the percentage of negative $\boldsymbol{\beta}$ compared to the non-shaded positive $\boldsymbol{\beta}$. HO-DGBN models are constructed using the original time series (complete), the stationary temporal signal, the non-stationary, and the time-series residuals. Each HO-DGBN is also constructed using a daily time series, along with a three-day block mean, weekly, biweekly and monthly means. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR) and Chlorophyll-a (Chl-a).

6.5.7 Lake Model Coefficients: Monthly – Complete Time Series

The non-stationary and residual HO-DGBN returned significantly higher error than their complete and stationary counterparts. These models enabled forecasting of $chl - a_0$ via time-series reconstruction at a higher accuracy compared to the complete HO-DGBN at the daily-biweekly Δ. However, monthly HO-DGBN at the monthly scale had outperformed the reconstructed for some lakes and returned forecasting errors on par with the stationary data. To better understand how parameter interactions differ at the lake level, only the complete time series at the monthly Δ is used (Figure 6.15). The complete time series includes any non-stationary trends, which are important in algal biomass dynamics. Due to its performance and the importance of the non-stationary trend (despite violation of the assumption of stationarity), further assessments of interaction effects between parameters and $chl - a_0$ at the lake level are conducted for only HO-DGBN at the monthly scale using the complete time series.

Across all lakes, the majority of SNSR, SR, T2m and past observations of chl-a parent nodes returned a positive β ($\geq$ 60% of parent nodes), while LMLD, WS and PPT returned primarily negative β ($\leq$ 40% of parent nodes). LSWT and SSR parent nodes returned a mixed β (40 – 60% of parent nodes), varying by lake and t. When separated by lake regions, NL showed a higher proportion of positive LSWT and T2m parent nodes compared to SL, where the effects were mixed. Additionally, NL saw a greater proportion of negative WS and SR parent nodes compared to SL (Table 6.6).

Table 6.6. Proportion of HO-DGBN parent nodes that exhibited a positive or negative coeffecient (β) across all lakes, only Northern Lakes (Great Bear Lake, Great Slave Lake, and Lake Athabasca), and only Southern Lakes (Lake Winnipeg and Lake Erie), for each input parameter. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR) and Chlorophyll-a (Chl-a).

Region	All Lakes		Northern Lakes		Southern Lakes	
Direction	Positive	Negative	Positive	Negative	Positive	Negative
LMLD	40.00%	60.00%	46.15%	53.85%	33.33%	66.67%
LSWT	57.14%	42.86%	64.71%	35.29%	45.45%	54.55%
SNSR	71.43%	28.57%	61.54%	38.46%	87.50%	12.50%
SR	70.83%	29.17%	64.71%	35.29%	85.71%	14.29%
SSR	56.00%	44.00%	53.33%	46.67%	60.00%	40.00%
T2m	70.00%	30.00%	83.33%	16.67%	50.00%	50.00%

Region	All Lakes		Northern Lakes		Southern Lakes	
Direction	Positive	Negative	Positive	Negative	Positive	Negative
WS	30.77%	69.23%	16.67%	83.33%	42.86%	57.14%
PPT	37.50%	62.50%	37.50%	62.50%	37.50%	62.50%
Chl-*a*	63.64%	36.36%	72.22%	27.78%	53.33%	46.67%
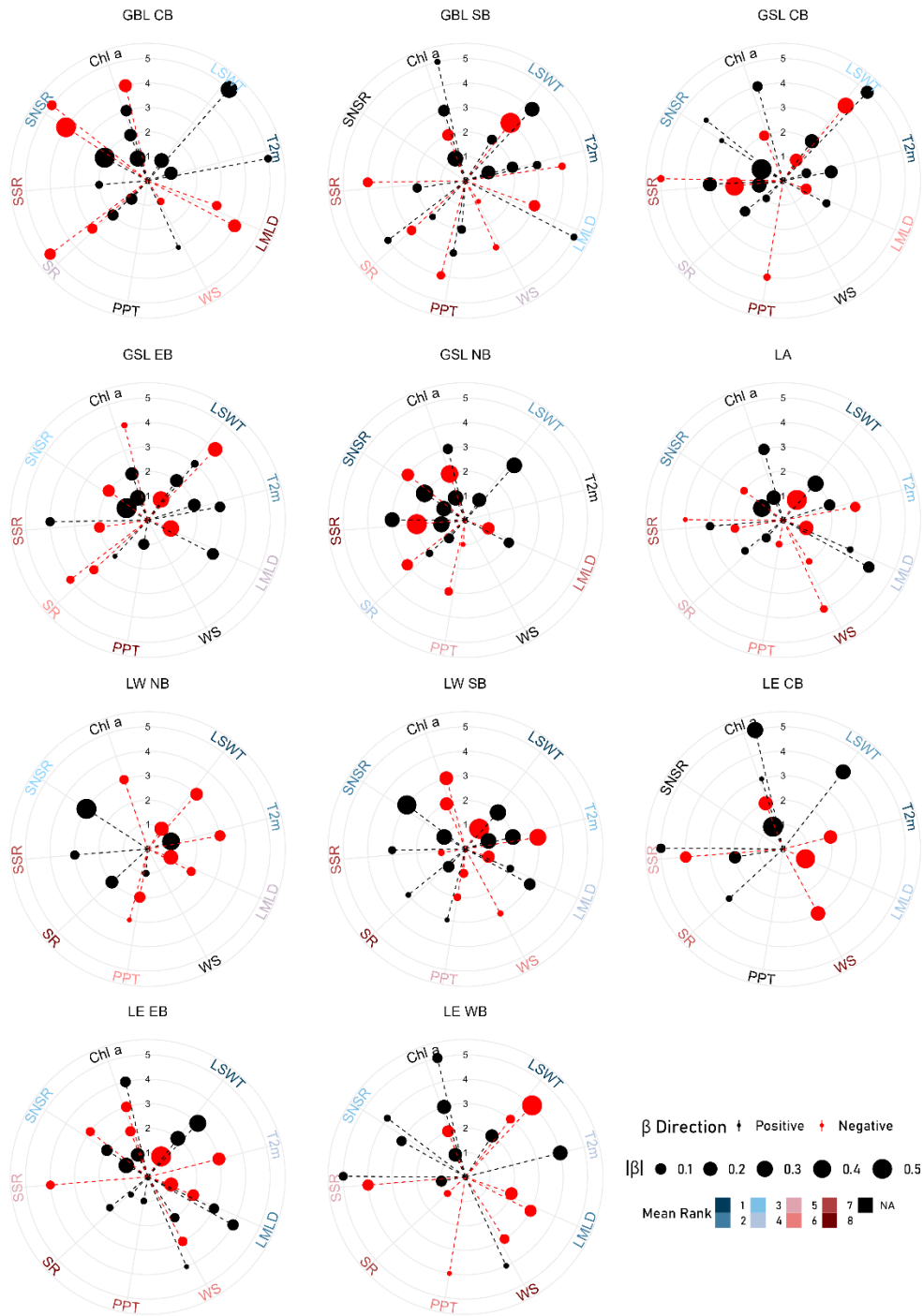
Figure 6.15. HO-DGBN mapped network using monthly block means and the original time-series (complete). The colours of the dots represent the model coeffecients ($\boldsymbol{\beta}$) direction (positive vs. negative). The size of the dots represents the magnitude of the slope in absolute values. Colours for the labels represent the derived parameter importance (Figure 6.12), where 1 is highest importance, 8 is lowest, and NA indicates the HO-DGBN model has no parent nodes of that parameter (in

174

the case of chl-a, it cannot be excluded as the target parameter, therefore no importance sampling was done). The y-axis represents the time-steps. HO-DGBN models are constructed for 5 North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin and WB = West Basin. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR), Surface Runoff (SR) and Chlorophyll-a (Chl-a).

Of GBL, the CB returned no PPT parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, LSWT, T2m and SSR returned positive interactions at all t, while LMLD returned all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with SNSR (34.52%) and chl-a (22.27%). $SNSR_2$ showed a positive effect on $chl - a_0$ while $SNSR_{4-5}$ showed a negative effect. $chl - a_{1-3}$ showed a positive effect on $chl - a_0$, while $chl - a_4$ showed a negative effect. SR returned the largest relative β_{rj} contribution (summed across all t) in GBL CB (11.83%) compared to all other lakes. Within the SB of GBL, SNSR returned no parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, WS was the only parameter which exhibited an entirely negative interaction at all t, while all other parameters returned mixed effects. The majority of β_{rj} contribution occurred with LSWT (39.59%) and chl-a (24.43%). LSWT returned a shifting positive-to-negative effect on $chl - a_0$ as t increased, while $chl - a_{1,3,5}$ returned a positive effect and $chl - a_2$ returned a negative effect.

Of GSL, the CB returned no WS parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, T2m, SNSR, and SR returned positive interactions at all t, while PPT returned all negative. Past observations of chl-a, LSWT, LMLD, and SSR all returned mixed effects depending on t. The majority of β_{rj} contribution occurred with SSR (31.16%) and LSWT (28.13%). SSR returned a shifting positive-to-negative effect on $chl - a_0$ as t increased, while LSWT returned a shifting negative-to-positive effect with increasing t. SSR returned the largest relative β_{rj} contribution (summed across all t) in GSL CB compared to all other lakes. Within the EB, WS had also returned no parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, T2m and PPT returned positive interactions at all t, while all other parameters exhibited mixed effects. The majority of β_{rj} contribution occurred with LSWT (25.53%) and SNSR (23.58%). $LSWT_{1,4}$ showed a positive

effect on $chl - a_0$ while $LSWT_{2,3}$ showed a negative effect. $SNSR_1$ showed a positive effect on $chl - a_0$ while $SNSR_2$ showed a negative effect. Within the NB, both T2m and WS exhibited no parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, LSWT returned positive interactions at all t, while PPT returned all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with SSR (30.60%) and SNSR (22.08%). SSR returned a shifting positive-to-negative effect on $chl - a_0$ as t increased. $SNSR_{1,2}$ showed a positive effect on $chl - a_0$ while $SNSR_3$ showed a negative effect.

Of LA, SR and Chl-a returned an entirely positive interaction at all t, while WS and PPT returned ally negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with LSWT (36.26%). LSWT returned a shifting negative-to-positive effect on $chl - a_0$ as t increased.

Of LW, the NB returned no WS parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, SR, SSR and SNSR returned positive interactions at all t, while past observations of chl-a, LSWT and LMLD returned all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with SNSR (26.46%) and T2m (23.80%). T2m returned a shifting positive-to-negative effect on $chl - a_0$ as t increased. Within the SB, SR and SNSR returned positive interactions at all t, while past observations of chl-a, and WS returned all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with T2m (26.88%), LSWT (25.39%) and SNSR (22.45%). LSWT returned a shifting negative-to-positive effect on $chl - a_0$ as t increased, while $T2m_{1,2}$ returned a positive effect, with $T2m_3$ returning a negative effect.

Of LE, the CB returned no PPT or SNSR parent nodes for $chl - a_0$ at any given t. Of the parameters included in the HO-DGBN, LSWT and SR returned positive interactions at all t, while T2m, LMLD and WS returned all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with past observations of

176

chl-a (42.45%). chl $- a_{1,3,5}$ returned a positive effect on chl $- a_0$, while chl $- a_2$ returned a negative effect. Within the EB, PPT, and SR returned positive interactions at all t, while SSR was all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with LSWT (41.01%), which ranked 1st in variable importance (Figure 6.12). $LSWT_1$ returned a negative effect on chl $- a_0$, while $LSWT_{2,3}$ returned a positive effect. Within the WB, T2m and SNSR returned positive interactions at all t, while LMLD, PPT and SR returned all negative. All other parameters returned mixed effects depending on t. The majority of β_{rj} contribution occurred with LSWT (34.02%) and past observations of chl-a (26.08%). $LSWT_2$ returned a positive effect on chl $- a_0$, while $LSWT_{3,4}$ returned a negative effect. chl $- a_{1,3,5}$ returned a positive effect on chl $- a_0$, while chl $- a_2$ returned a negative effect.

6.5.8 Model Sensitivity: Monthly – Complete Time Series

The sensitivity of each HO-DGBN determines how much the forecasted $chl - a_0$ varies from the original forecast given a percent change in the normalized parameter data. Parameters with a high MAD indicated high sensitivity to change in that given parameter. Across all lakes, SNSR returned the highest median MAD for the Monthly – Complete HO-DGBN (0.72 μg L^{-1}), followed by LSWT, T2m, LMLD, WS, SR, PPT and SSR (Table 6.7). LSWT returned the highest MAD for the largest number of lakes (63.4%; 7/11), while SNSR returned the highest MAD in GBL SB and GSL NB, and T2m returned the highest MAD in GBL CB and LW SB. Although LMLD returned minimal sensitivity in most lakes, all basins of LE showed a similar relative MAD with LSWT, T2m and SNSR compared to other parameters (0.28 – 4.00 μg L^{-1}). While LSWT, T2m and SNSR returned the highest overall MAD across all lakes, the shift in forecasted $chl - a_0$ varies by lake throughout the projected year.

Table 6.7. Sensitivity results where each normalized input parameter was iteratively adjusted by -30%, -15%, -5%, 5% 15% and 30%, where the Mean Absolute Deviation (MAD) is calculated from the adjusted forecasted chl-a and from the original forecasted chl-a concentrations using HO-DGBN constructed using monthly block means of the original time-series (complete). HO-DGBN models are constructed for 5 North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin and WB = West Basin. Parameters include Lake Lake Surface Water

Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR) and Surface Runoff (SR). All values are represent as µg L^{-1}.

	LSWT	T2m	PPT	WS	LMLD	SNSR	SSR	SR
GBL CB	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01
GBL SB	0.28	0.21	0.07	0.14	0.19	0.33	0.03	0.12
GSL CB	0.34	0.15	0.02	0.10	0.20	0.20	0.01	0.07
GSL EB	0.22	0.22	0.04	0.09	0.14	0.21	0.03	0.04
GSL NB	0.55	0.39	0.07	0.09	0.25	0.95	0.05	0.08
LA AB	0.29	0.20	0.11	0.09	0.17	0.14	0.05	0.10
LW NB	1.18	0.80	0.12	0.19	0.30	0.92	0.06	0.08
LW SB	0.54	1.20	0.17	0.15	0.24	0.83	0.12	0.18
LE CB	9.86	1.53	0.33	0.72	4.00	8.19	0.46	0.13
LE EB	0.93	0.72	0.04	0.52	0.28	0.74	0.07	0.06
LE WB	2.09	1.51	0.04	0.26	0.50	0.72	0.24	0.07
Median	**0.54**	**0.39**	**0.07**	**0.14**	**0.24**	**0.72**	**0.05**	**0.08**

When LSWT was increased, GBL and GSL saw a decrease in spring/early summer $chl - a_0$ and an increase in the late summer/fall. LA saw multiple shifts, with a decrease in early spring $chl - a_0$, increase in late spring, decrease in summer, followed by an increase in fall. LW saw an increase in spring $chl - a_0$, and similar to LA, returned a decrease in summer, and an increase in fall. Finally, LE saw a decrease in spring $chl - a_0$, which then increased in the early summer, followed by decrease in late summer, and a slight increase fall.

When T2m was increased, GBL, GSL, and LA all saw a decrease in spring $chl - a_0$, and an increase in the summer/fall. Similarly, LW NB saw a decrease in spring $chl - a_0$ and increase in summer but saw a decrease in the fall. LW SB and LE saw a different pattern compared to other lakes, with an increase in spring $chl - a_0$, a decrease in early summer, followed by an increase in late summer/fall.

When SNSR was increased, NL saw minimal change in forecasted $chl - a_0$, however, the SL saw a significant change. LW NB saw an increase in spring – fall $chl - a_0$, while the SB saw a decrease in spring, a high increase in the summer, and little change in the fall. The response in LE varied by basin. LE CB saw an increase in the spring $chl - a_0$, which decreased

in the summer, and increased again in the fall. LE WB returned a similar trend, with an increase in early spring $chl - a_0$, a decrease in late spring/early summer, which then increased in late summer/fall. Finally, LE EB saw a decrease in spring $chl - a_0$, which then increase in the summer, and showed little change in the fall.

When LMLD increased, most lakes saw minimal change in forecasted $chl - a_0$, however LE saw a significant change. LE CB and EB saw an increase in spring $chl - a_0$, a decrease in the summer, and an increase in late summer/fall. LE WB returned similar results, except the fall $chl - a_0$ decreased with the increase in LMLD. All other parameters showed minimal effect on the forecasted chl-a. Sensitivity results time series can be seen in Figure 6.16.
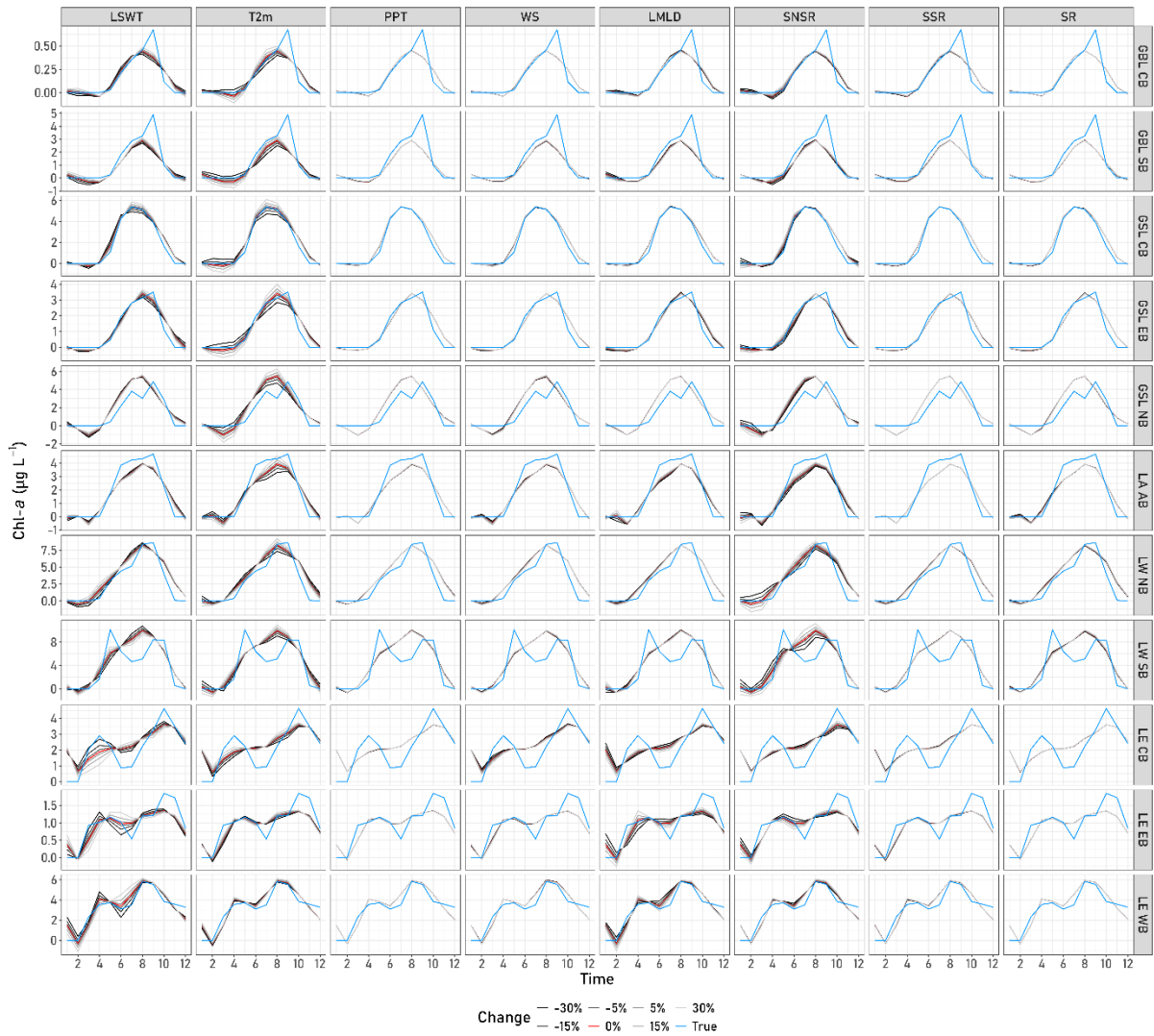
Figure 6.16. Sensitivity results of the forecasted chl-a concentrations results where each normalized input parameter was iteratively adjusted by -30%, -15%, -5%, 5% 15%. The blue line represents the satellite-derived chl-a concentrations, while the red line represents the original HO-DGBN chl-a forecast. HO-DGBN models are constructed for 5 North American Great Lakes and their basins, where GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie, CB = Central Basin, SB = South Basin, EB = East Basin, NB = North Basin and WB = West Basin. Parameters include Lake Lake Surface Water Temperature (LSWT), 2m Air Temperature (T2m), Total Precipitation (PPT), Wind Speed (WS), Lake Mixing Level Depth (LMLD), Surface Net Solar Radiation (SNSR), Sub Surface Runoff (SSR) and Surface Runoff (SR).

6.6 Discussion

6.6.1 Time Series

HO-DGBN allow for a potential breakdown of complex systems to a more understandable level, where parameter interactions are described as the density function of a linear Gaussian regression, which is directed toward a given parameter. This method assumes stationarity and homogeneity across time, which rarely occurs within real world data (Hoekstra, *et al.*, 2012). Despite this assumption, several studies have implemented DBNs with non-stationary data (Quesada *et al.*, 2021a; Robinson *et al.*, 2010). Most input parameters had significant trends (Figure 6.6), however the decreased frequency of significance with increasing Δ was likely a result of a reduced sample size. The increase in Theil-Sen slope increasing Δ may have been a result of the minimization of variance from temporal means, which may also reduce the influence of overwintering stagnation (*i.e.* the assumption of 0.00 $\mu g\ L^{-1}$ for chl-*a* or 273.15 K for LSWT, and even reduced/prolonged overwinter runoff). The increase in significant negative LMLD trends for NL indicates a potential increase to stratification, likely a result of the significantly higher increase in T2m for the region, compared to the SL (Hansen *et al.*, 2006; Houze *et al.*, 2019). Conversely, the increase in LMLD for LE was potentially caused by the increase in PPT, inducing more frequent mixing events (Holgerson *et al.*, 2022), including the spring season (Williams & King, 2020). Due to the presence of significant trends, the time series were decomposed to the stationary, non-stationary and residual signals, to determine if the use of a complete time-series significantly impacts the quality of the network.

6.6.2 Overview of the HO-DGBN

The use of high order Markovian models provided higher quality networks with lower forecasting error (Figure 6.7). This may be due to the delayed response that several parameters may have in providing direct effects on algal biomass dynamics. A high order was found to be impactful regardless of Δ. However, both the stationary and complete HO-DGBN models showed a significant reduction in error with increasing Δ. The reduction in error with

increasing Δ may be a result of the minimization of variance derived from the remote sensing product. Per pixel accuracy in retrieved chl-*a* may return a wide range of values, where deriving a single daily value from a spatial mean reduced the variance. However, daily fluctuations persist (Figure 6.2), particularly during the ice on/off period. A temporal average helps minimize the variance seen during these periods, returning lower error. It is also possible that by increasing Δ, there is a stronger interaction effect between parameter observations weeks – months prior and $chl - a_0$, than those only up to 1-5 days. Reconstructing the forecasted chl-*a* derived from stationary, non-stationary and residual HO-DGBN did reduce the total error compared to the use of a complete time series; however, this was mainly true at a lower Δ, where the difference at the monthly Δ was less significant. The Thel-Sein slopes were greater at the monthly scale, however, there were little to no significant Mann-Kendall, therefore the time-series at the monthly Δ is primarily stationary. Models consisting of stationary data return stronger networks, as the non-stationary and residual inputs performed poorly, and yielded little to no change in forecasting error with the change in Δ. The poor performance of the HO-DGBN at a smaller Δ, and for the non-stationary and residual models limit our confidence in the network structure accurately explaining algal biomass dynamics. While the inputs for the daily Δ have higher variance compared to the monthly Δ, the magnitude of forecasted chl-*a* concentrations remains lower (Figure 6.9) compared to that of the monthly data (Figure 6.10).

6.6.3 Interaction Effects Between Parameters and Chl-*a*

Despite the poor performance of the shorter Δ and non-stationary/residual models, some interesting results were observed. Importance sampling imposes a mutilated network with a new set of data observations (evidence), by excluding a given parameter, where the network excludes the parent nodes of that parameter and reweights the model (Koller & Friedman, 2009). As $chl - a_0$ is the predictor target, it cannot be excluded from the network.

It was found that LSWT was the most important parameter across all Δ, similar to other studies that consistently observe LSWT as one of the most important parameters in algal biomass dynamics (Borowitzka *et al.*, 2016; Paerl & Huisman, 2008). It is often assumed that

182

higher LSWT equates to higher algal biomass; however, this is not always the case. LSWT may exceed the optima for the given algal community, slowing algal growth (Paerl & Otten, 2013). Warmer waters will also increase the metabolic rate of algae, where nutrient-poor waters may exhibit a negative effect on algal growth (Kraemer *et al.*, 2017). For the monthly – complete HO-DGBN (Figure 6.15), both oligotrophic (*e.g*., GSL EB) and eutrophic lake basins (*e.g.*, LE WB), returned both positive and negative effects with LSWT. Each parent node time step is at a monthly interval; therefore, periods of high LSWT may be associated with nutrient depletion due to higher algal metabolic rates during warmer periods. Therefore, a period of increasing LSWT may have a negative impact on $chl - a_0$. This is also true of past observations of chl-*a*, where months prior to increasing algae may return a negative effect on $chl - a_0$.

The increase in T2m and SNSR showed greater importance at an increased Δ. The relationship between LSWT and T2m is not always linear and can vary between lakes (Rose *et al.*, 2016; Walsh *et al.*, 2020). Due to the high heat capacity of water, changes in T2m take a longer time to affect the LSWT; therefore, values derived from a higher Δ may create a stronger network for T2m. A negative interaction between T2m and $chl - a_0$ is always at $t > 1$, as there is often a delay between increasing LSWT and raising T2m; therefore, the potential depletion of nutrients impact may be further delayed compared to LSWT. The HO-DGBN models in NL were found to return a higher proportion of positive parent nodes for LSWT and T2m compared to SL (Table 6.6). Therefore, temperature has a more consistent positive effect on chl-*a* concentrations in NL, where temperatures return a higher rate of change. SNSR also exhibits a lag period between the onset of increased SNSR and algal biomass growth, lasting days to potentially weeks due to other factors such as water column stability, temperature, salinity, *etc.* (Williams, 1967; Sforza *et al.*, 2012; Persuad *et al.*, 2015). The importance of SNSR peaked at the biweekly Δ, and still showed high importance at a smaller Δ with the complete time-series. Light facilitates algal growth, where higher intensity can increase the metabolic rate of algae (Maltsev *et al.*, 2021). Too high of a light intensity may also lower algal growth and cell count, as seen in culture studies (Sforza *et al.*, 2012). SNSR exhibited a

higher proportion of negative β_{rj} contribution in the NL (Figure 6.14b) compared to the SL (Figure 6.14c), along with a higher proportion of negative parent nodes (Table 6.6) at the monthly Δ, likely a result of higher nutrient concentrations in the SL.

SR and SSR association with algal biomass typically aid in the mobilization of nutrients such as DOC, P and N (Garnier *et al.*, 2022; Ali & English, 2019). However, SR exhibited low importance compared to other parameters, while SSR only exhibited high importance for some SL at the daily Δ. This indicates that SSR has a more immediate influence on algal biomass dynamics compared to other parameters, where Figure 6.14 showed that this interaction was predominantly negative. While the importance of SSR was considered low in the NL, Figure 6.14b indicates a very high contribution to the networks total β_{rj}, particularly at the daily Δ, driven mainly by LA. The NL watersheds contain sporadic discontinuous permafrost, extensive continuous permafrost, and continuous permafrost (Sapriza-azuri *et al.*, 2018). Permafrost regions may experience less or channelized SSR, potentially limiting its importance for those regions (Andresen *et al.*, 2020). However, with thawing permafrost this may change, as can be seen in Figure 6.12, where the non-stationary signal returned high SSR importance at the daily scale.

The LMLD returned higher importance at an increased Δ for the stationary HO-DGBN, where a shallow LMLD would indicate primarily stagnate waters, and high LMLD would indicate mixing. Mixing disperses surface water algae, while increasing nutrient availability throughout the water column (Reynolds, 1994). LMLD exhibits two peaks a year, a smaller one during early spring season, and a larger one during the fall (Figure 6.5). As the LMLD increases in the spring, the chl-a began to increase, however the inverse is true in the fall (Figure 6.5). At the monthly Δ, when chl-a is at its peak, past observations of LMLD are at a minimum, providing an association between previously stable conditions with higher algal biomass yield. This association is clearer when using a monthly Δ as the temporal lag reaches as far back as five months. The effect of LMLD on $chl - a_0$ was predominantly negative across all lakes, with a higher percentage for SL (Table 6.6). As seen in Figure 6.15, no lakes returned a positive interaction between LMLD and $chl - a_0$ at $t = 1$, however positive

interactions do occur at $t > 1$, indicating a typical immediate negative effect. PPT and WS both returned a low importance and contributed little to the total β_{rj} of the networks. PPT is expected to not only contribute to external loading (Coffey *et al.*, 2019; Creed & Band 1998), but also contribute to lake mixing (Ho & Michalak, 2020), similar to the effects of WS (Huang et al 2016; Zhou et al 2021). Due to the size and depth of the study lakes, PPT and WS are unlikely to provide a significant impact, as wind-driven mixing and flushing are more impactful in shallower waterbodies (Kämpf, 2017). PPT does show some importance with the stationary HO-DGBN at the daily – three-day Δ (Figure 6.12), particularly for the shallower GBL SB, LA and LW.

6.6.4 HO-DGBN Sensitivity

Sensitivity analysis provides insight into how sensitive the HO-DGBN model is to a change in each parameter, and how chl-a forecasts might change given a shift in these parameters. Unsurprisingly, variables with low importance, or low β_{rj} contribution, show little impact on the forecast of $chl - a_0$ when changed. Instead, variables such LSWT, T2m, SNSR and to an extent LMLD provided some influential change to the forecasted $chl - a_0$. Despite the uniform change in each parameter, the forecasted $chl - a_0$ did not change uniformly. For the NL, the peak algal biomass had increased, along with chl-a concentrations later in the season (Figure 6.16). T2m also increased the algal biomass peak; however, it also increased in the preceding months compared to the latter of the LSWT. The NL exhibit shorter summer stratification periods and, as oligotrophic systems, return a singular algal biomass peak (Sommer *et al.*, 2012). Therefore, the increase in temperature is expected to increase this peak. However, rising temperatures and earlier ice off timings are also anticipated to affect the rate of stratification, where earlier stratification may lead to an additional spring biomass peak (Sommer *et al.*, 2012). The SL differed with respect to the impacts of temperature, where LW indicated increased spring algal biomass, while LE showed an increase in early summer algal biomass (Figure 6.16). However, T2m returned an inverse relationship, where the summer algal biomass peaks increased in LW, and both the spring and fall algal biomass peaks increased in LE. High spring temperatures leading to earlier onset of stratification, could also

result in incomplete mixing, and limiting the duration of internal loading, potentially limiting early algal biomass growth (Favot *et al.*, 2019). However, earlier onset of stratification is associated with higher biomass later in the season (Favot *et al.*, 2019). Additional research into developing future climate scenarios on lake system effects is possible with the provided networks. The application of DGBNs to scenario analysis has been done in the past (Molina *et al.*, 2013; Xie *et al.*, 2022) and is highly applicable to this research.

6.7 Conclusions

This research showcases the application of the ESA CCI Lakes product in identifying interactions between daily chl-*a* observations and various lake physical and atmospheric parameters. HO-DGBN provide a robust method in identifying complex interactions between parameters as they change over time. It was found that temperature (LSWT and T2m) was the most important parameter and the largest contributor in the network to changes in algal biomass; however, these interactions were not always positive. Temperature parameter interactions with algal biomasses were more consistently positive for NL, compared to SL. Other important parameters such as SNSR generally had a positive effect on chl-*a* concentrations in all lakes, while LMLD exhibited a negative interaction. Factors such as PPT and WS returned little effect on algal biomass, while SR and SSR returned a higher importance at a shorter Δ. HO-DGBN models returned higher forecasting accuracy when constructed with monthly averaged parameters compared to the daily scale. The accuracy of the networks depends on the accuracy of the inputs, where future updates to the ESA CCI Lakes product that improve the chl-*a* retrieval accuracy and fill missing data gaps can further improve the HO-DGBN models, particularly at a low Δ.

Chapter 7: Conclusions

7.1 Summary

The primary objective of this thesis was to investigate the impact of lake physical and atmospheric variables on the trends, variability, and dynamics of algal biomass in five North American Great Lakes determined from the ESA CCI Lakes and ECMWF (ERA5-land hourly) data products available for the period 2002-2020 . This work is vital, as described in Chapter 1, due to the ever present and emerging threat algal blooms pose to freshwater resources, impacting environmental, economic, and human health. Climate change is anticipated to drive the rise in algal blooms, threatening even remote, oligotrophic lakes. However, few comprehensive observational analyses have been conducted due to *in situ* data limitations. This research provided a novel approach to the application of remote sensing data products in analyzing algal biomass dynamics and interactions with various lake physical and atmospheric forcings.

Chapters 2 provided an extensive overview of how climate change affects lake phenology, the dynamics of algal growth, community composition impacts, how remote sensing techniques are used to retrieve water quality parameters, and how various statistical techniques may be used to determine interaction effects amongst multiple variables. Within Chapter 2, it was shown that changes are occurring in abiotic factors such as air temperatures, LSWT, lake mixing regimes, LIC duration and extent, stratification, precipitation (duration, frequency and severity), wind speeds and CO2 depositions, and how these factors impact algal biomass growth, retention, dissipation and community composition. Anthropogenic activities have contributed to greater mobilization of allochthonous macro and micronutrients that stimulate algal growth, increase legacy P, and alter surface runoff of NAPs. Algae can be monitored through the use of optical satellite imagery, where empirical, machine learning, and bio-optical methods can make use of various wavelength channels to parse the chl-*a* signal from other NAPs and organic matter. Chapter 3 provided an overview of information regarding the general study area and lakes, where the Northern Lakes (NL; GBL, GSL and LA) are considerably deeper, cooler and oligotrophic compared to the warmer, shallower and eutrophic

Southern Lakes (SL; LW and LE). This research made use of the ESA CCI Lakes product (v2.0.0), providing daily LWLR, LSWT and LIC data on a ca. 1 km gridded. Additional 9 km gridded data of lake physical and atmospheric parameters (T2m, WS, PPT, LMLD, SNSR, SR and SSR) were provided by the ERA5-Land hourly product.

Due to the lack of LIC studies on the analysis of algal biomass trends, Chapter 4 provided assessment of the ESA CCI Lakes data product for the five target lakes (2002 – 2020). A total of five algal biomass parameters (mean & max chl-*a*, abnormal chl-*a*, abnormal algal biomass days, and peak chl-*a* DOY), four LSWT (mean & max LSWT, anomaly LSWT days, and peak LSWT DOY), and three LIC (Ice on/off and duration) parameters were constructed annually and seasonally (spring, summer, and fall). It was found that ice-off periods changed by a total of -0.55 – 0.43 days^{-1} yr^{-1}, annual maximum LSWTs by -0.03 – 0.11 K^{-1} yr^{-1} and mean annual chl-*a* concentrations by -0.02 – 0.17 µg L^{-1} yr^{-1}. Vector Autoregression models with a lag order of three years were developed using all combinations annual and seasonal (spring, summer and fall) of LIC and LSWT parameters to determine which provided a greater explanation of variance in algal biomass over time. It was found that NLs (GBL, GSL and LA) had a higher percentage of variance explained by LIC parameters on average, compared to SLs (LW and LE). SLs provided a poor explanation for spring algal biomass parameters, regardless of the predictive parameter. On average, NLs provided a greater explanation of the variance for chl-*a* concentrations (mean adj. r^2 = 0.59-0.79) compared to SLs (mean adj r^2 = 0.36-0.38). The results of this study provide evidence that; (i) climate parameters provide greater explanatory power for remote NLs, and (ii) LIC has a more significant explanation of variance for algal biomass phenology compared to more eutrophic and anthropogenically impacted lakes.

The results of Chapter 4 indicated that LIC had a significant impact on the phenology of algal biomass, particularly for NLs. Chapter 5 expanded on this concept to determine whether LIC may act as a predictor of the parameters of algal biomass. Using the annual and seasonal parameters developed in Chapter 4, MLR and ANN models were constructed to determine whether the LIC parameters significantly contribute to the prediction of algal

biomass compared to the LSWT parameters. ANN models (median NRMSE = 0.81) significantly outperformed MLR models (median NRMSE = 0.95) in predicting algal biomass parameters. Models using only LSWT parameters returned significantly lower error (median NRMSE = 0.82) compared to LIC only models (median NRMSE = 0.93), while models including both LSWT and LIC returned the lowest error (median NRRMSE = 0.75). A random forest (RF) model was constructed to classify abnormal algal biomass presence for each lake/basin and season with varying accuracies (72.88% − 97.27%, median = 84.22%). It was found that the LIC parameters were the most important on average in classifying the presence of abnormal algal biomass presence within NLs (GBL, GSL, LA). LIC was not found to be a strong predictor of algal biomass parameters; however, it was an important variable in classifying abnormal springtime algal biomass concentrations, primarily in NLs. Due to the control of LIC on the timing and duration of mixing, it is a significant contributor to algal biomass growth in lakes which rely primarily on internal loading for nutrient availability.

Chapter 5 provided evidence for the importance of LIC in remote NLs for the springtime algal biomass. However, LIC parameters can only assess the annual/seasonal trends and variance of algal biomass, thereby missing the intra-year trends, cycles, and variance. Additionally, various lake physical and atmospheric parameters were not included. Chapter 6 saw the construction of various HO-DGBN to identify the interaction effects between various atmospheric (T2m, PPT, WS & SNSR) and lake physical (LSWT, LMLD, SR and SSR) parameters with chl-a, as they change over time. HO-DGBN models were constructed for all lakes in the study area, using the original time series, stationary, non-stationary and residual signals, in varying temporal ranges (Δ; daily, three days, weekly, biweekly and monthly averages). It was found that a HO-DGBN constructed using the stationary signal returned the lowest forecasting error (NRMSE = 0.08 − 0.71, median = 0.31), however the original time-series using the monthly Δ returned similar error (NRMSE = 0.16 − 0.55, median = 0.39). Through importance sampling, LSWT and T2m returned the highest error when excluded from the HO-DGBN (median MAE = 1.70 & 1.60 µg L^{-1} respectively), followed closely by SNSR (median MAE = 1.41 µg L^{-1}). Temperature interactions were mixed, typically returning both a

positive and negative interaction depending on the HO-DGBN time-slice and lake. LMLD returned increased importance at a longer Δ, while SSR returned increased importance at a shorter Δ. SNSR typically exhibited a positive interaction with chl-*a*, while LMLD exhibited a frequent negative interaction. The PPT and WS were found to be the least important parameters across all the study lakes. New remote sensing data products provide significant potential for network analysis using observational data, where future scenario analysis could provide improved forecasting in algal biomass dynamics in a changing climate.

7.2 Limitations

The primary limiting factor for this research is the accuracy of the ESA CCI Lakes product, as accurately retrieving chl-*a* concentrations across many optical water types, geolocations, and satellite sensors is highly difficult. As discussed in section 3.3, there are many complications which can convolute or overlap the chl-*a* signal within a turbid water body, often a result of NAPs, varying atmospheric conditions, and alterations in the illumination angles. It was found that observations from 2012-2015 were missing with this product, eliminating the potential to understand some dynamics within this period, which would improve the HO-DGBN models. Additionally, it was found that the transition from MERIS (2002 – 2020) to Sentinel-3A (2016 – present), returned a slight bias for some lakes, where Sentinel-3A returned higher values compared to that of MERIS. However, it is unknown if this is a sensor/retrieval bias, or due to a process of increasing algal biomass as seen in many lakes (Kraemer *et al.,* 2022).

7.3 Future Work

The methods presented in this thesis showcase the potential of remote sensing data products to better understand algal biomass dynamics. By filling in missing data, it is possible to identify intricate trends and patterns, which are often unique to a given lake basin. Remote sensing products still return missing features, and while this thesis utilized methods to interpolate missing data, it was not the primary focus of the research. Improved data interpolation through non-linear means may improve the accuracy of the time-series.

Applications to a greater range of lakes would also provide greater context and validity to statements made between the Northern and Southern lake regions. Additionally, improvements to remote sensing data products would require a reapplication of the methods presented in this thesis to update the results and improve their accuracy. Additional environmental parameters such as land use dynamics (*e.g.*, reduction of natural forest or wetland environments), atmospheric deposition of CO_2, NO_x and NH_3, and macro and micronutrients were not included in this thesis, where future research may look to find methods of integrating these types of data. The purpose of improving our understanding of the dynamics of algal biomass in relation to changing atmospheric and lake physical variables is to improve forecasting and projection models. Integration of the findings presented in this thesis into future projection forecasts is a potential direction for future work in this field.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. 12th Symposium on Operating Systems Design and Implementation ({OSDI} 16) (pp. 265-283).

Adrian, R., Hessen, D. O., Blenckner, T., Hillebrand, H., Hilt, S., Jeppesen, E. ... & Trolle, D. (2016). Environmental impacts—Lake ecosystems. In North Sea Region Climate Change Assessment (pp. 315-340). Springer, Cham. https://doi.org/10.1007/978-3-319-39745-0_10

Adrian, R., Walz, N., Hintze, T., Hoeg, S., & Rusche, R. (1999). Effects of Winter Conditions on the Plankton Succession During Spring in a Shallow Polymictic Lake. Freshwater Biology, 41, 621–632. https://doi.org/10.1046/j.1365-2427.1999.00411.x

Adrian, R., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., Livingstone, D. M. … Van Donk, E. (2009). Lakes as Sentinels of Climate Change. Limnology and Oceanography, 54: 2283-2297. Limnology and Oceanography, 54(6), 2283.

Afrin, T., & Yodo, N. (2021). A Probabilistic Estimation of Traffic Congestion Using Bayesian Network. Measurement, 174, 109051. https://doi.org/10.1016/j.measurement.2021.109051

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., & Salmerón, A. (2011). Bayesian Networks in Environmental Modelling. Environmental Modelling & Software, 26(12), 1376-1388. https://doi.org/10.1016/j.envsoft.2011.06.004

Ahn, C. Y., Chung, A. S., & Oh, H. M. (2002). Rainfall, Phycocyanin, and N:P Ratios Related to Cyanobacterial Blooms in a Korean Large Reservoir. Hydrobiologia, 474, 117–124. https://doi.org/10.1023/A:1016573225220

Ai, H., Zhang, K., Sun, J., & Zhang, H. (2023). Short-term Lake Erie Algal Bloom Prediction by Classification and Regression Models. Water Research, 232, 119710–119710. https://doi.org/10.1016/j.watres.2023.119710

Ajaj, Pradhan, B., Noori, A. M., & Jebur, M. N. (2017). Spatial Monitoring of Desertification Extent in Western Iraq using Landsat Images and GIS. Land Degradation & Development, 28(8), 2418–2431. https://doi.org/10.1002/ldr.2775

Ali, & English, C. (2019). Phytoplankton Blooms in Lake Winnipeg Linked to Selective Water-Gatekeeper Connectivity. Scientific Reports, 9(1), 8395–8395. https://doi.org/10.1038/s41598-019-44717-y

Allan, M. G., Hamilton, D. P., Hicks, B., & Brabyn, L. (2015). Empirical and Semi-Analytical Chlorophyll-a Algorithms for Multi-Temporal Monitoring of New Zealand Lakes using Landsat. Environmental Monitoring and Assessment, 187(6). https://doi.org/10.1007/s10661-015-4585-4

Amemiya, A., Shlok, M., & Miyoshi, T. (2023). Application of Recurrent Neural Networks to Model Bias Correction: Idealized Experiments with the Lorenz-96 Model. Journal of Advances in Modeling Earth Systems, 15(2), e2022MS003164. https://doi.org/10.1029/2022MS00316

Anderson, J. T., Larsen, R. K., & Lacouture, R. V. (2010). Assessing the Magnitude of Polycyclic Aromatic Hydrocarbon Loading from Road Surfaces and its Effect on Algal Productivity. Last viewed: 04/21/2023 at: https://rosap.ntl.bts.gov/view/dot/34445

Andresen, C. G., Lawrence, D. M., Wilson, C. J., McGuire, A. D., Koven, C., Schaefer, K., Jafarov, E., Peng, S., Chen, X., Gouttevin, I., Burke, E., Chadburn, S., Ji, D., Chen, G., Hayes, D., & Zhang, W. (2020). Soil Moisture and Hydrology Projections of the Permafrost Region – a Model Intercomparison. The Cryosphere, 14(2), 445–459. https://doi.org/10.5194/tc-14-445-2020

Apicella, A., Donnarumma, F., Isgrò, F., & Prevete, R. (2021). A Survey on Modern Trainable Activation Functions. Neural Networks, 138, 14-32. https://doi.org/10.1016/j.neunet.2021.01.026

Ariano, S. S., & Brown, L. C. (2019). Ice Processes on Medium-Wized North-Temperate Lakes. Hydrological Processes, 33(18), 2434–2448. https://doi.org/10.1002/hyp.13481

Arif, C., Setiawan, B. I., & Hasanah, N. A. I. (2020, September). Predicting Methane Emission from Paddy Fields with Limited Soil Data by Artificial Neural Networks. In 2020 International Conference on Computer Science and its Application in Agriculture (ICOSICA) (pp. 1-4). IEEE. https://doi.org/10.1109/ICOSICA49951.2020.9243192

Atkins. W. R. G. (1924). On the Vertical Mixing of Sea-Water and its Importance for the Algal Plankton. Journal of the Marine Biological Association of the United Kingdom, 13(2), 319–324. https://doi.org/10.1017/S0025315400007530

Avalon Rare Metals Inc. (2011). Thor Lake Project: Great Slave Lake Section 2.8.6. Avalon Rare Metals Incorporated, as submitted to the Mackenzie Valley Land and Water Board. Last accessed: 4/21/2022 at https://reviewboard.ca/upload/project_document/EA1011-001_02_Thor_Lake_Project_DAR_Description_of_Existing_Biophysical_Environment__206-230_.PDF

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support Vector Regression. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, 67-80.

Ayala-Borda, P., Lovejoy, C., Power, M., & Rautio, M. (2021). Evidence of Eutrophication in Arctic Lakes. Arctic Science, 7(4), 859-871. https://doi.org/10.1139/as-2020-0033

Bade, D. L., Carpenter, S. R., Cole, J. J., Pace, M. L., Kritzberg, E., Van De Bogert, M. C., … McKnight, D. M. (2007). Sources and Fates of Dissolved Organic Carbon in Lakes

as Determined by Whole-Lake Carbon Isotope Additions. Biogeochemistry, 84(2), 115–129. https://doi.org/10.1007/s10533-006-9013-y

Bakker, E. S., & Hilt, S. (2016). Impact of Water-Level Fluctuations on Cyanobacterial Blooms: Options for Management. Aquatic Ecology, 50(3), 485–498. https://doi.org/10.1007/s10452-015-9556-x

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., & Betts, A. K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. Journal of hydrometeorology, 10(3), 623-643. https://doi.org/10.1175/2008JHM1068.1

Barton, D. N., Kuikka, S., Varis, O., Uusitalo, L., Henriksen, H. J., Borsuk, M., ... & Linnell, J. D. (2012). Bayesian Networks in Environmental and Resource Management. Integrated Environmental Assessment and Management, 8(3), 418-429. https://doi.org/10.1002/ieam.1327

Bartosiewicz, M., Przytulska, A., Deshpande, B. N., Antoniades, D., Cortes, A., MacIntyre, S., Lehmann, M. F., & Laurion, I. (2019). Effects of Climate Change and Episodic Heat Events on Cyanobacteria in a Eutrophic Polymictic Lake. The Science of the Total Environment, 693, 133414–. https://doi.org/10.1016/j.scitotenv.2019.07.220

Baruah, P. J., Tamura, M., Oki, K., & Nishimura, H. (2001). Neural Network Modeling of Lake Surface Chlorophyll and Sediment Content from Landsat TM imagery. 22nd Asian Conference on Remote Sensing, (November), 5–9. Retrieved from http://www.crisp.nus.edu.sig/~acrs2001/

Baruah, P. J., Tamura, M., Oki, K., & Nishimura, H. (2002). Neural Network Modeling of Surface Chlorophyll and Sediment Content in Inland Water from Landsat Thematic Mapper Imagery Using Multidate Spectrometer Data. Ocean Optics: Remote Sensing and Underwater Imaging, 4488, 205–212. https://doi.org/10.1117/12.452815

Batur, E., & Maktav, D. (2019). Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCA Method in the Lake Gala, Turkey. IEEE Transactions on Geoscience and Remote Sensing, 57(5), 2983–2989. https://doi.org/10.1109/TGRS.2018.2879024

Bégin, P. N., Tanabe, Y., Kumagai, M., Culley, A. I., Paquette, M., Sarrazin, D., ... & Vincent, W. F. (2021). Extreme Warming and Regime Shift Toward Amplified Variability in a Far Northern Lake. Limnology and Oceanography, 66, S17-S29. https://doi.org/10.1002/lno.11546

Bennett, R. M., Card, J. R., & Hornby, D. M. (1973). Hydrology of Lake Athabasca: Past, Present and Future. Hydrological Sciences Journal, 18(3), 337-345. https://doi.org/10.1080/02626667309494045

Benson, B. J., Magnuson, J. J., Jensen, O. P., Card, V. M., Hodgkins, G., Korhonen, J., … Granin, N. G. (2012). Extreme Events, Trends, and Variability in Northern Hemisphere

Lake-Ice Phenology (1855-2005). Climatic Change, 112(2), 299–323. https://doi.org/10.1007/s10584-011-0212-8

Beretta, S., Castelli, M., Gonçalves, I., Henriques, R., & Ramazzotti, D. (2018). Learning the Structure of Bayesian Networks: a Quantitative Assessment of the Effect of Different Algorithmic Schemes. Complexity, 1–12. https://doi.org/10.1155/2018/1591878

Betts, A. K., Reid, D., & Crossett, C. (2020). Evaluation of the FLake Model in ERA5 for Lake Champlain. Frontiers in Environmental Science, 8, 250. https://doi.org/10.3389/fenvs.2020.609254

Beyene, M. T., & Jain, S. (2020). Climate-Related Thresholds in Lake Ice and the Associated Environmental and Social Systems. Water Security, 10, 100065–. https://doi.org/10.1016/j.wasec.2020.100065

Bielski, A., & Toś, C. (2022). Remote Sensing of the Water Quality Parameters for a Shallow Dam Reservoir. Applied Sciences, 12(13), 6734. https://doi.org/10.3390/app12136734

Bill, L., T. Flett, I. Unka, E. Beaver, L. Mercredi, S. Martin, and G. McDonald (1994). Northern River Basins Study Traditional Knowledge Documentation Project. Edmonton: Government of Alberta.

Binding, C. E., Greenberg, T. ., McCullough, G., Watson, S. ., & Page, E. (2018). An Analysis of Satellite-Derived Chlorophyll and Algal Bloom Indices on Lake Winnipeg. Journal of Great Lakes Research, 44(3), 436–446. https://doi.org/10.1016/j.jglr.2018.04.001

Binding, C. E., Pizzolato, L., & Zeng, C. (2021). EOLakeWatch; Delivering a Comprehensive Suite of Remote Sensing Algal Bloom Indices for Enhanced Monitoring of Canadian Eutrophic Lakes. Ecological Indicators, 121, 106999. https://doi.org/10.1016/j.ecolind.2020.106999

Binding, C. E., Zastepa, A., & Zeng, C. (2019). The Impact of Phytoplankton Community Composition on Optical Properties and Satellite Observations of the 2017 Western Lake Erie Algal Bloom. Journal of Great Lakes Research, 45(3), 573-586. https://doi.org/10.1016/j.jglr.2018.04.001

Blenckner, T., Adrian, R., Livingstone, D. M., Jennings, E., Weyhenmeyer, G. A., George, D. G., … Teubner, K. (2007). Large-Scale Climatic Signatures in Lakes Across Europe: a Meta-Analysis. Global Change Biology, 13(7), 1314–1326. https://doi.org/10.1111/j.1365-2486.2007.01364.x

Block, B. D., Denfeld, B. A., Stockwell, J. D., Flaim, G., Grossart, H. P. F., Knoll, L. B., … Hampton, S. E. (2019). The Unique Methodological Challenges of Winter Limnology. Limnology and Oceanography: Methods, 17(1), 42–57. https://doi.org/10.1002/lom3.10295

Boehrer, B., & Schultze, M. (2008). Stratification of lakes. Reviews of Geophysics, 46(2). https://doi.org/10.1029/2006RG000210

Bonachela, J. A., Raghib, M., & Levin, S. A. (2012). Erratum: Dynamic Model of Flexible Phytoplankton Nutrient Uptake (Proceedings of the National Academy of Sciences of the United States of America.108, 51, (20633-20638) https://doi.org/10.1073/pnas.1118012108).

Borowitzka, M. A., Beardall, J., & Raven, J. A. (Eds.). (2016). The Physiology of Microalgae (Vol. 6). Cham: Springer.

Bosse, K. R., Sayers, M. J., Shuchman, R. A., Fahnenstiel, G. L., Ruberg, S. A., Fanslow, D. L., … Burtner, A. M. (2019). Spatial-Temporal Variability of in situ Cyanobacteria Vertical Structure in Western Lake Erie: Implications for Remote Sensing Observations. Journal of Great Lakes Research, 45(3), 480–489. https://doi.org/10.1016/j.jglr.2019.02.003

Bøttcher, S. G., & Dethlefsen, C. (2003). deal: A Package for Learning Bayesian Networks. Journal of Statistical Software, 8, 1-40. https://doi.org/10.18637/jss.v008.i20

Brasil, J., Attayde, J. L., Vasconcelos, F. R., Dantas, D. D. F., & Huszar, V. L. M. (2016). Drought-Induced Water-Level Reduction Favors Cyanobacteria Blooms in Tropical Shallow Lakes. Hydrobiologia, 770(1), 145–164. https://doi.org/10.1007/s10750-015-2578-5

Breiman, L. (1996). Bagging Predictors. Machine Learning, 24, 123-140.

Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and Regression Trees. CRC press.

Brezonik, P., Menken, K. D., & Bauer, M. (2005). Landsat-Based Remote Sensing of Lake Water Quality Characteristics, Including Chlorophyll and Colored Dissolved Organic Matter (CDOM). Lake and Reservoir Management, 21(4), 373–382. https://doi.org/10.1080/07438140509354442

Brooks, B. W., Lazorchak, J. M., Howard, M. D. A., Johnson, M.-V. V., Morton, S. L., Perkins, D. A. K., Reavie, E. D., Scott, G. I., Smith, S. A., & Steevens, J. A. (2016). Are Harmful Algal Blooms Becoming the Greatest Inland Water Quality Threat to Public Health and Aquatic Ecosystems? Environmental Toxicology and Chemistry, 35(1), 6–13. https://doi.org/10.1002/etc.3220

Brown, L. C., & Duguay, C. R. (2010). The Response and Role of Ice Cover in Lake-Climate Interactions. Progress in Physical Geography, 34(5), 671–704. https://doi.org/10.1177/0309133310375653

Budd, J. W., Beeton, A. M., Stumpf, R. P., Culver, D. A., & Charles Kerfoot, W. (2001). Satellite Observations of Microcystis Blooms in Western Lake Erie. Internationale

Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen, 27(7), 3787-3793. https://doi.org/10.1080/03680770.1998.11901692

Bukata, R. P., Jerome, J. H., Kondratyev, A. S., & Pozdnyakov, D. V. (1995). Optical Properties and Remote Sensing of Inland and Coastal Waters (1st ed.). Taylor & Francis Group. https://doi.org/10.1201/9780203744956

Butcher, J. B., Nover, D., Johnson, T. E., & Clark, C. M. (2015). Sensitivity of Lake Thermal and Mixing Dynamics to Climate Change. Climatic Change, 129(1–2), 295–305. https://doi.org/10.1007/s10584-015-1326-1

Cai, Y., Duguay, C. R., & Ke, C.-Q. (2022). A 41-Year (1979–2019) Passive-Microwave-Derived Lake Ice Phenology Data Record of the Northern Hemisphere. Earth System Science Data, 14(7), 3329–3347. https://doi.org/10.5194/essd-14-3329-2022

Cameron, E. S., Muller, K. M., Stone, M., Buttle, J., Leach, J., Webster, K., & Emelko, M. B. (2022). Early Seasonal Increases and Persistence in Relative Abundance of Potentially Toxic Cyanobacteria: Concerning Impacts of Extended Ice-Free Periods in Northern Temperate Lakes. bioRxiv, 2022-12. https://doi.org/10.1101/2022.12.20.521158

Carder, K. L., Chen, F. R., Cannizzaro, J. P., Campbell, J. W., & Mitchell, B. G. (2004). Performance of the MODIS Semi-Analytical Ocean Color Algorithm For Chlorophyll-a. Advances in Space Research, 33(7), 1152–1159. https://doi.org/10.1016/S0273-1177(03)00365-X

Carder, K. L., Chen, F. R., Lee, Z. P., Hawes, S. K., & Kamykowski, D. (1999). Semianalytic Moderate-Resolution Imaging Spectrometer Algorithms for Chlorophyll-a and Absorption with Bio-Optical Domains Based on Nitrate-Depletion Temperatures. Journal of Geophysical Research: Oceans, 104(C3), 5403-5421. https://doi.org/10.1029/1998JC900082

Carmichael, W. (2008). A World Overview--One-Hundred-Twenty-Seven Years of Research on Toxic Cyanobacteria - Where do we go from here? Advances in Experimental Medicine and Biology, 619, 104–125. https://doi.org/10.1007/978-0-387-75865-7_4

Caron, D. A., & Hutchins, D. A. (2013). The Effects of Changing Climate on Microzooplankton Grazing and Community Structure: Drivers, Predictions and Knowledge Gaps. Journal of Plankton Research, 35(2), 235–252. https://doi.org/10.1093/plankt/fbs091

Carr, N.G., Whitton B.A., (1982). The Biology of Cyanobacteria. Botanical Monographs, Vol. 19, University of California Press, Blackwell Scientific Publications, Blackwell, Oxford.

Carrea, L., J.-F. Crétaux, X. Liu, Y. Wu, M. Bergé-Nguyen, B. Calmettes, C.R. Duguay, C.J. Merchant, N. Selmes, S.G. H. Simis, M. Warren, H. Yesou, D. Müller, D. Jiang, and C. Albergel (2023). Satellite-Derived Multivariate World-Wide Lake Physical Variable

Timeseries for Climate Studies. Scientific Data, 10:30, https://doi.org/10.1038/s41597-022-01889-z .

Carrea, L.; Crétaux, J.-F.; Liu, X.; Wu, Y.; Bergé-Nguyen, M.; Calmettes, B.; Duguay, C.; Jiang, D.; Merchant, C.J.; Mueller, D.; Selmes, N.; Spyrakos, E.; Simis, S.; Stelzer, K.; Warren, M.; Yesou, H.; Zhang, D. (2022): ESA Lakes Climate Change Initiative (Lakes_cci): Lake products, Version 2.0. NERC EDS Centre for Environmental Data Analysis, 18 March 2022. doi:10.5285/ab8d21568c81491fbb9a300c36884af7. http://dx.doi.org/10.5285/ab8d21568c81491fbb9a300c36884af7

Carrick, H. J., Worth, D., & Marshall, M. L. (1994). The Influence of Water Circulation on Chlorophyll-Turbidity Relationships in Lake Okeechobee as Determined by Remote Sensing. Journal of Plankton Research, 16(9), 1117–1135. https://doi.org/10.1093/plankt/16.9.1117

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., & Westermann, S. (2017). An Observation-Based Constraint on Permafrost Loss as a Function of Global Warming. Nature Climate Change, 7(5), 340–344. https://doi.org/10.1038/nclimate3262

Chang, J., Bai, Y., Xue, J., Gong, L., Zeng, F., Sun, H., ... & Ma, Y. (2023). Dynamic Bayesian Networks with Application in Environmental Modeling and Management: A Review. Environmental Modelling & Software, 105835. https://doi.org/10.1016/j.envsoft.2023.105835

Chang, N. B., & Bai, K. (2018). Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing. CRC Press. Boca Raton, FL, US.

Charfi, S., BenHamad, S., & Masmoudi, A. (2020). Assessing the Impact of Monetary Fundamentals on Exchange Rate Fluctuations A Bayesian Network Approach. Journal of Modelling in Management, 15(1), 166-181. https://doi.org/10.1108/JM2-09-2018-0130

Chavarie, Howland, K., Harris, L., & Tonn, W. (2015). Polymorphism in Lake Trout in Great Bear Lake: Intra-Lake Morphological Diversification at Two Spatial Scales. Biological Journal of the Linnean Society, 114(1), 109–125. https://doi.org/10.1111/bij.12398

Chavez, P. S. (1988). An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data. Remote Sensing of Environment, 24(3), 459–479. https://doi.org/10.1016/0034-4257(88)90019-3

Chavez, P. S. (1996). Image-Based Atmospheric Corrections-Revisited and Improved. Photogrammetric Engineering and Remote Sensing, 62(9), 1025-1035.

Chen, A., Ren, Z., Guo, W., Liang, Y., & Feng, Z. (2023). An Efficient Adaptive Differential Grouping Algorithm for Large-Scale Black-Box Optimization. IEEE Transactions on Evolutionary Computation, 27(3), 1–1. https://doi.org/10.1109/TEVC.2022.3170793

Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., ... & Hoek, G. (2019b). A Comparison of Linear Regression, Regularization, And Machine Learning Algorithms to Develop Europe-Wide Spatial Models of Fine Particles and Nitrogen Dioxide. Environment International, 130, 104934. https://doi.org/10.1016/j.envint.2019.104934

Chen, J., Zhu, W., Tian, Y. Q., Yu, Q., Zheng, Y., & Huang, L. (2017). Remote Estimation of Colored Dissolved Organic Matter and Chlorophyll-a in Lake Huron Using Sentinel-2 Measurements. Journal of Applied Remote Sensing, 11(03), 1. https://doi.org/10.1117/1.jrs.11.036007

Chen, M., Wang, D., Ding, S., Fan, X., Jin, Z., Wu, Y., … Zhang, C. (2019a). Zinc Pollution in Zones Dominated by Algae and Submerged Macrophytes in Lake Taihu. Science of the Total Environment, 670, 361–368. https://doi.org/10.1016/j.scitotenv.2019.03.167

Chételat, J., Cott, P. A., Rosabal, M., Houben, A., McClelland, C., Rose, E. B., & Amyot, M. (2019). Arsenic Bioaccumulation in Subarctic Fishes of a Mine-Impacted Bay on Great Slave Lake, Northwest Territories, Canada. PloS One, 14(8), e0221361–. https://doi.org/10.1371/journal.pone.0221361

Chibwe, L., Muir, D. C. G., Gopalapillai, Y., Shang, D., Kirk, J. L., Manzano, C. A., Atkinson, B., Wang, X., & Teixeira, C. (2021). Long-term spatial and temporal trends, and source apportionment of polycyclic aromatic compounds in the Athabasca Oil Sands Region. *Environmental Pollution (1987)*, *268*(Pt A), 115351-. https://doi.org/10.1016/j.envpol.2020.115351

Chinain, M., Germain, M., Deparis, X., Pauillac, S., & Legrand, A. M. (1999). Seasonal Abundance and Toxicity of the Dinoflagellate Gambierdiscus Spp. (Dinophyceae), The Causative Agent of Ciguatera in Tahiti, French Polynesia. Marine Biology, 135(2), 259–267. https://doi.org/10.1007/s002270050623

Coats, R., Perez-Losada, J., Schladow, G., Richards, R., & Goldman, C. (2006). The Warming of Lake Tahoe. Climatic Change, 76(1–2), 121–148. https://doi.org/10.1007/s10584-005-9006-1

Coffey, R., Paul, M. J., Stamp, J., Hamilton, A., & Johnson, T. (2019). A Review of Water Quality Responses to Air Temperature and Precipitation Changes 2: Nutrients, Algal Blooms, Sediment, Pathogens. Journal of the American Water Resources Association, 55(4), 844–868. https://doi.org/10.1111/1752-1688.12711

Cole J. J. & Prairie Y. T. (2009) Dissolved CO2. In: Gene E. Likens, (Editor) Encyclopedia of Inland Waters. Volume 2, pp. 30-34 Oxford: Elsevier.

Coleman, K. A., Hoskin, G. N., Chasmer, L., Thienpont, J. R., Quinton, W. L., & Korosi, J. B. (2023). Limnology and Diatom Ecology of Shallow Lakes in a Rapidly Thawing Discontinuous Permafrost Peatland. Inland Waters (Print), 13(1), 13–29. https://doi.org/10.1080/20442041.2022.2144699

Commission for Environmental Cooperation (CEC) (2007). North America Elevation 1-Kilometer Resolution GRID. Derived from the GTOPO30 Elevation Dataset. U.S.Geological Survey and the Center for Topographic Information (Sherbrooke), Natural Resources Canada. Last Accessed; 09/01/2023 at: https://www.sciencebase.gov/catalog/item/4fb5495ee4b04cb937751d6d

Conley, D., Paerl, H., Howarth, R., Boesch, D., Seitzinger, S., Havens, K., Lancelot, C., & Likens, G. (2009). Controlling Eutrophication: Nitrogen and Phosphorus. Science. 323. 1014-1015. https://doi.org/10.1126/science.1167755

Creed, I. F., & Band, L. E. (1998). Export Of Nitrogen from Catchments within a Temperate Forest: Evidence for a Unifying Mechanism Regulated by Variable Source Area Dynamics. Water Resources Research, 34(11), 3105–3120. https://doi.org/10.1029/98WR01924

Creed, I. F., Bergström, A. K., Trick, C. G., Grimm, N. B., Hessen, D. O., Karlsson, J., … Weyhenmeyer, G. A. (2018). Global Change-Driven Effects on Dissolved Organic Matter Composition: Implications for Food Webs of Northern Lakes. Global Change Biology, 24(8), 3692–3714. https://doi.org/10.1111/gcb.14129

Cross, W. F., Hood, J. M., Benstead, J. P., Huryn, A. D., & Nelson, D. (2015). Interactions Between Temperature and Nutrients Across Levels of Ecological Organization. Global Change Biology, 21(3), 1025–1040. https://doi.org/10.1111/gcb.12809

Crossman, J., Futter, M. N., Elliott, J. A., Whitehead, P. G., Jin, L., & Dillon, P. J. (2019). Optimizing Land Management Strategies for Maximum Improvements in Lake Dissolved Oxygen Concentrations. Science of the Total Environment, 652, 382–397. https://doi.org/10.1016/j.scitotenv.2018.10.160

Cruz, R.C., Costa, P. R., Vinga, S., Krippahl, L., & Lopes, M. B. (2021). A Review of Recent Machine Learning Advances for Forecasting Harmful Algal Blooms and Shellfish Contamination. Journal of Marine Science and Engineering, 9(3), 283–. https://doi.org/10.3390/jmse9030283

D'Ambrosio, S. L., & Harrison, J. A. (2021). Methanogenesis Exceeds CH4 Consumption in Eutrophic Lake Sediments. Limnology and Oceanography Letters, 6(4), 173–181. https://doi.org/10.1002/lol2.10192

Dabrowski, J. J., Beyers, C., & de Villiers, J. P. (2016). Systemic Banking Crisis Early Warning Systems Using Dynamic Bayesian Networks. Expert Systems with Applications, 62, 225-242. https://doi.org/10.1016/j.eswa.2016.06.024

Dai, A., Zhao, T., & Chen, J. (2018). Climate Change and Drought: A Precipitation and Evaporation Perspective. Current Climate Change Reports, 4(3), 301–312. https://doi.org/10.1007/s40641-018-0101-6

Dallosch, M. A., & Creed, I. F. (2021). Optimization of Landsat Chl-a Retrieval Algorithms in Freshwater Lakes through Classification of Optical Water Types. Remote Sensing (Basel, Switzerland), 13(22), 4607–. https://doi.org/10.3390/rs13224607

Dalton, C., Sparber, K., & de Eyto, E. (2018). Assessing Sedimentation in a Temperate Dystrophic Lake in the NE Atlantic Seaboard Region. Journal of Paleolimnology, 60(2), 117–131. https://doi.org/10.1007/s10933-018-0022-3

Dang, X., Du, J., Wang, C., Zhang, F., Wu, L., Liu, J., ... & Wang, J. (2023). A Hybrid Chlorophyll-a Estimation Method for Oligotrophic and Mesotrophic Reservoirs Based on Optical Water Classification. Remote Sensing, 15(8), 2209. https://doi.org/10.3390/rs15082209

De Keukelaere, L., Moelans, R., Knaeps, E., Sterckx, S., Reusen, I., De Munck, D., ... & Tyler, A. (2023). Airborne Drones for Water Quality Mapping in Inland, Transitional and Coastal Waters—MapEO Water Data Processing and Validation. Remote Sensing, 15(5), 1345.

De Moraes Novo, E. M. L., de Farias Barbosa, C. C., Freitas, R. M., Shimabukuro, Y. E., Melack, J. M., & Filho, W. P. (2006). Seasonal Changes in Chlorophyll Distributions in Amazon Floodplain Lakes Derived from MODIS Images. Limnology, 7(3), 153–161. https://doi.org/10.1007/s10201-006-0179-8

De Senerpont Domis, L. N., Mooij, W. M., & Huisman, J. (2007). Climate-Induced Shifts in an Experimental Phytoplankton Community: A Mechanistic Approach. Hydrobiologia, 584(1), 403–413. https://doi.org/10.1007/s10750-007-0609-6

Deng, J., Qin, B., Paerl, H. W., Zhang, Y., Ma, J., & Chen, Y. (2014). Earlier and Warmer Springs Increase Cyanobacterial (Microcystis Spp.) Blooms in Subtropical Lake Taihu, China. Freshwater Biology, 59(5), 1076-1085. https://doi.org/10.1111/fwb.12330

Deng, J., Shan, K., Shi, K., Qian, S. S., Zhang, Y., Qin, B., & Zhu, G. (2023). Nutrient Reduction Mitigated the Expansion of Cyanobacterial Blooms Caused by Climate Change in Lake Taihu According to Bayesian Network Models. Water Research, 236, 119946. https://doi.org/10.1016/j.watres.2023.119946

Devlin, M. J., Barry, J., Mills, D. K., Gowen, R. J., Foden, J., Sivyer, D., & Tett, P. (2008). Relationships Between Suspended Particulate Material, Light Attenuation and Secchi Depth in UK Marine Waters. Estuarine, Coastal and Shelf Science, 79(3), 429-439. https://doi.org/10.1016/j.ecss.2008.04.024

Dibike, Y., Prowse, T., Saloranta, T., & Ahmed, R. (2011). Response of Northern Hemisphere Lake-Ice Cover And Lake-Water Thermal Structure Patterns to a Changing Climate. Hydrological Processes, 25(19), 2942-2953. https://doi.org/10.1002/hyp.8068

Dinnat, E. P., & Le Vine, D. M. (2008). Impact of Sun Glint on Salinity Remote Sensing: An Example with the Aquarius Radiometer. IEEE Transactions on Geoscience and Remote Sensing, 46(10), 3137–3150. https://doi.org/10.1109/TGRS.2008.2000629

Docker, M. F., Bravener, G. A., Garroway, C. J., Hrodey, P. J., Hume, J. B., Johnson, N. S., ... & Zollweg-Horan, E. C. (2021). A Review of Sea Lamprey Dispersal and Population Structure in the Great Lakes and the Implications for Control. Journal of Great Lakes Research, 47, S549-S569. https://doi.org/10.1016/j.jglr.2021.09.015

Domingos, P. (2012). A Few Useful Things to Know About Machine Learning. Communications of the ACM, 55(10), 78-87. https://doi.org/10.1145/2347736.2347755

Dong, L., Du, H., Han, N., Li, X., Zhu, D., Mao, F., … He, S. (2020). Application of Convolutional Neural Network on Lei Bamboo Above-Ground-Biomass (AGB) Estimation Using Worldview-2. Remote Sensing, 12(6). https://doi.org/10.3390/rs12060958

Du, J., Kimball, J. S., Duguay, C., Kim, Y., & Watts, J. D. (2017). Satellite Microwave Assessment of Northern Hemisphere Lake Ice Phenology from 2002 to 2015. Cryosphere, 11(1), 47–63. https://doi.org/10.5194/tc-11-47-2017

Dugan. H. A. (2021). A Comparison of Ecological Memory of Lake Ice-Off in Eight North-Temperate Lakes. Journal of Geophysical Research. Biogeosciences, 126(6). https://doi.org/10.1029/2020JG006232

ECMWF (2016). IFS Documentation – Cy43r1 Operational implementation 22 Nov 2016, Part IV: Physical Processes. European Centre for Medium-Range Weather Forecasts. Last Accessed: 4/22/2022 at https://www.ecmwf.int/sites/default/files/elibrary/2016/17117-part-iv-physical-processes.pdf#subsection.H.6.3

Edlund, Ramstack Hobbs, J. M., Heathcote, A. J., Engstrom, D. R., Saros, J. E., Strock, K. E., Hobbs, W. O., Andresen, N. A., & VanderMeulen, D. D. (2022). Physical Characteristics of Northern Forested Lakes Predict Sensitivity to Climate Change. Hydrobiologia, 849(12), 2705–2729. https://doi.org/10.1007/s10750-022-04887-9

Eleveld, M. A., Ruescas, A. B., Hommersom, A., Moore, T. S., Peters, S. W., & Brockmann, C. (2017). An Optical Classification Tool for Global Lake Waters. Remote Sensing, 9(5), 420. https://doi.org/10.3390/rs9050420

Emadodin, I., Reinsch, T., & Taube, F. (2019). Drought and Desertification in Iran. Hydrology, 6(3). https://doi.org/10.3390/hydrology6030066

Environmental Protection Agency (EPA). (2020). The Great Lakes: Lake Erie. Environmental Protection Agency (EPA). Last Accessed: 04/20/2022 at https://www.epa.gov/greatlakes/lake-erie

Erratt, K. J., Creed, I. F., & Trick, C. G. (2018). Comparative Effects Of Ammonium, Nitrate and Urea on Growth And Photosynthetic Efficiency of Three Bloom-Forming Cyanobacteria. Freshwater Biology, 63(7), 626–638. https://doi.org/10.1111/fwb.13099

Evans. (2000). The large lake ecosystems of Northern Canada. Aquatic Ecosystem Health & Management, 3(1), 65–79. https://doi.org/10.1080/14634980008656992

Evans, M. S., & Muir, D. C. G. (2016). Persistent organic contaminants in sediments and biota of Great Slave Lake, Canada: Slave River and long-range atmospheric source influences. *Journal of Great Lakes Research*, *42*(2), 233–247. https://doi.org/10.1016/j.jglr.2015.12.001

Falbel, D., Allaire, J.J., Chollet, F., Tang, Y., Van Der Bijl, W., Studer., Keydana, S. (2019). R Interface to Keras. Github. Found at: https://github.com/rstudio/keras/

Falkowski, P. G., & Raven, J. A. (2007). Aquatic Photosynthesis. Princeton University Press.

Fang, X., & Sommer, U. (2017). Overwintering Effects on the Spring Bloom Dynamics of Phytoplankton. Journal of Plankton Research, 39(5), 772–780. https://doi.org/10.1093/plankt/fbx046

Favot, E. J., Holeton, C., DeSellas, A. M., & Paterson, A. M. (2023). Cyanobacterial Blooms in Ontario, Canada: Continued Increase in Reports Through the 21st Century. Lake and Reservoir Management, 39(1), 1–20. https://doi.org/10.1080/10402381.2022.2157781

Favot, E. J., Rühland, K. M., DeSellas, A. M., Ingram, R., Paterson, A. M., & Smol, J. P. (2019). Climate Variability Promotes Unprecedented Cyanobacterial Blooms in a Remote, Oligotrophic Ontario Lake: Evidence From Paleolimnology. Journal of Paleolimnology, 62(1), 31-52. https://doi.org/10.1007/s10933-019-00074-4

Feiner, Z. S., Dugan, H. A., Lottig, N. R., Sass, G. G., & Gerrish, G. A. (2022). A Perspective on the Ecological and Evolutionary Consequences of Phenological Variability in Lake Ice on North-Temperate Lakes. Canadian Journal of Fisheries and Aquatic Sciences, 79(9), 1590–1604. https://doi.org/10.1139/cjfas-2021-0221

Ficke, A. D., Myrick, C. A., & Hansen, L. J. (2007). Potential Impacts of Global Climate Change on Freshwater Fisheries. In Reviews in Fish Biology and Fisheries (Vol. 17). https://doi.org/10.1007/s11160-007-9059-5

Ficker, H., Luger, M., & Gassner, H. (2017). From Dimictic to Monomictic: Empirical Evidence of Thermal Regime Transitions in Three Deep Alpine Lakes in Austria Induced by Climate Change. Freshwater Biology, 62(8), 1335–1345. https://doi.org/10.1111/fwb.12946

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. Science, 281(5374), 237–240. https://doi.org/10.1126/science.281.5374.237

Filatov, N., Baklagin, V., Efremova, T., Nazarova, L., & Palshin, N. (2019). Climate Change Impacts on the Watersheds of Lakes Onego and Ladoga from Remote Sensing and in-situ Data. Inland Waters (Print), 9(2), 130–141. https://doi.org/10.1080/20442041.2018.1533355

Finney, B. P., Rühland, K., Smol, J. P., & Fallu, M. A. (2004). Paleolimnology of the North American Subarctic. Long-Term Environmental Change in Arctic and Antarctic Lakes, 269-318. https://doi.org/10.1007/978-1-4020-2126-8_10

Flor, M., Weiß, M., Selhorst, T., Müller-Graf, C., & Greiner, M. (2020). Comparison of Bayesian and Frequentist Methods for Prevalence Estimation Under Misclassification. BMC Public Health, 20(1), 1-10. https://doi.org/10.1186/s12889-020-09177-4

Freeman, E. C., Creed, I. F., Jones, B., & Bergström, A. K. (2020). Global Changes may be Promoting a Rise in Select Cyanobacteria in Nutrient-Poor Northern Lakes. Global Change Biology, 26(9), 4966-4987. https://doi.org/10.1111/gcb.15189

Gakstatter, J. H., Allum, M. O., Dominguez, S. E., & Crouse, M. R. (1978). A Survey of Phosphorus and Nitrogen Levels in Treated Municipal Wastewater. Journal of the Water Pollution Control Federation, 50(4), 718–722.

Gameiro, C., Zwolinski, J., & Brotas, V. (2011). Light Control on Phytoplankton Production in a Shallow and Turbid Estuarine System. Hydrobiologia, 669(1), 249–263. https://doi.org/10.1007/s10750-011-0695-3

Garnier, J., Billen, G., Tournebize, J., Barré, P., Mary, B., & Baudin, F. (2022). Storage or Loss of Soil Active Carbon in Cropland Soils: The Effect of Agricultural Practices and Hydrology. Geoderma, 407, 115538–. https://doi.org/10.1016/j.geoderma.2021.115538

Gemela, J. (2001). Financial Analysis Using Bayesian Networks. Applied Stochastic Models in Business and Industry, 17(1), 57-67. https://doi.org/10.1002/asmb.422

Gerace, A. D., Schott, J. R., & Nevins, R. (2013). Increased Potential to Monitor Water Quality in the Near-Shore Environment with Landsat's Next-Generation Satellite. Journal Of Applied Remote Sensing, 7(1), 073558–073558. https://doi.org/10.1117/1.JRS.7.073558

Ghunowa, K., Medeiros, A. S., & Bello, R. (2019). Hyperspectral Analysis of Algal Biomass in Northern Lakes, Churchill, MB, Canada. Arctic Science, 5(4), 240–256. https://doi.org/10.1139/as-2018-0030

Giardino, C., Brando, V. E., Gege, P., Pinnel, N., Hochberg, E., Knaeps, E., ... & Dekker, A. (2019). Imaging Spectrometry of Inland and Coastal Waters: State of the Art, Achievements and Perspectives. Surveys in Geophysics, 40, 401-429. https://doi.org/10.1007/s10712-018-9476-0

Gibson, Prowse, T. D., & Peters, D. L. (2006). Hydroclimatic Controls on Water Balance and Water Level Variability in Great Slave Lake. Hydrological Processes, 20(19), 4155–4172. https://doi.org/10.1002/hyp.6424

Giménez, E. M., Winkler, G., Hoffmeyer, M., & Ferreyra, G. A. (2018). Composition, Spatial Distribution, and Trophic Structure of the Zooplankton Community in San Jorge Gulf, Southwestern Atlantic Ocean. Oceanography (Washington, D.C.), 31(4), 154–163. https://doi.org/10.5670/oceanog.2018.418

Gitelson, A. A., Gao, B. C., Li, R. R., Berdnikov, S., & Saprygin, V. (2011). Estimation of Chlorophyll-a Concentration in Productive Turbid Waters Using a Hyperspectral Imager for the Coastal Ocean - The Azov Sea Case Study. Environmental Research Letters, 6(2). https://doi.org/10.1088/1748-9326/6/2/024023

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships Between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. Journal of Plant Physiology, 160, 271–282.

Gobler, C. J. (2020). Climate Change and Harmful Algal Blooms: Insights and Perspective. Harmful Algae, 91(December 2019), 101731. https://doi.org/10.1016/j.hal.2019.101731

Gons, H. J. (1999). Optical Teledetection Of Chlorophyll-a in Turbid Inland Waters. Environmental Science and Technology, 33(7), 1127–1132. https://doi.org/10.1021/es9809657

Gons, H. J., Rijkeboer, M., & Ruddick, K. G. (2002). A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of Inland and Coastal Waters. Journal of Plankton Research, 24(9), 947–951. https://doi.org/10.1093/plankt/24.9.947

Gordon, H. R., & McCluney, W. R. (1975). Estimation of the Depth of Sunlight Penetration in the Sea for Remote Sensing. Applied Optics, 14(2), 413. https://doi.org/10.1364/ao.14.000413

Gorham. P. R. (1964). Toxic Algae as a Public Health Hazard. Journal - American Water Works Association, 56(11), 1481–1488. https://doi.org/10.1002/j.1551-8833.1964.tb01355.x

Government of Manitoba. (2021). Lake Winnipeg. Government of Manitoba: Environment, Climate and Parks. Last Accessed: 4/20/2022 at https://www.gov.mb.ca/water/lakes-beaches-rivers/lake-winnipeg.html

Graafland, C. E., & Gutiérrez, J. M. (2022). Learning Complex Dependency Structure of Gene Regulatory Networks from High Dimensional Microarray Data with Gaussian Bayesian Networks. Scientific Reports, 12(1), 18704. https://doi.org/10.1038/s41598-022-21957-z

Griffith, A. W., & Gobler, C. J. (2020). Harmful algal blooms: A Climate Change Co-Stressor in Marine and Freshwater Ecosystems. Harmful Algae, 91, 101590–101590. https://doi.org/10.1016/j.hal.2019.03.008

Groisman, P. Y., Knight, R. W., Karl, T. R., Easterling, D. R., Sun, B., & Lawrimore, J. H. (2004). Contemporary Changes of the Hydrological Cycle over the Contiguous United States: Trends Derived From In-Situ Observations. Journal of Hydrometeorology, 5(1), 64–85. https://doi.org/10.1175/1525-7541(2004)005<0064:CCOTHC>2.0.CO;2

Gronchi, E., Jöhnk, K. D., Straile, D., Diehl, S., & Peeters, F. (2021). Local and Continental-Scale Controls of the Onset of Spring Phytoplankton Blooms: Conclusions From a Proxy-Based Model. Global Change Biology, 27(9), 1976–1990. https://doi.org/10.1111/gcb.15521

Grzegorczyk. M. (2016). A Non-Homogeneous Dynamic Bayesian Network with a Hidden Markov Model Dependency Structure Among the Temporal Data Points. Machine Learning, 102(2), 155–207. https://doi.org/10.1007/s10994-015-5503-2

Gu, G., & Adler, R. F. (2023). Observed Variability and Trends in Global Precipitation During 1979–2020. Climate Dynamics, 61(1-2), 131–150. https://doi.org/10.1007/s00382-022-06567-9

Gujer, W. (2008). Time Series Analysis. Systems Analysis for Water Technology, 361-396.

Ha, N. T. T., Thao, N. T. P., Koike, K., & Nhuan, M. T. (2017). Selecting The Best Band Ratio to Estimate Chlorophyll-a Concentration in a Tropical Freshwater Lake Using Sentinel 2A Images From a Case Study of Lake Ba Be (Northern Vietnam). ISPRS International Journal of Geo-Information, 6(9). https://doi.org/10.3390/ijgi6090290

Haakonsson, S., Rodríguez-Gallego, L., Somma, A., & Bonilla, S. (2017). Temperature and Precipitation Shape the Distribution of Harmful Cyanobacteria in Subtropical Lotic and Lentic Ecosystems. Science of the Total Environment, 609, 1132–1139. https://doi.org/10.1016/j.scitotenv.2017.07.067

Hall, R. I., Wolfe, B. B., Wiklund, J. A., Edwards, T. W. D., Farwell, A. J., & Dixon, D. G. (2012). Has Alberta oil sands development altered delivery of polycyclic aromatic compounds to the Peace-Athabasca Delta? *PloS One*, *7*(9), e46089-. https://doi.org/10.1371/journal.pone.0046089

Hallegraeff, G. M. (2010). Ocean Climate Change, Phytoplankton Community Responses, and Harmful Algal Blooms: A Formidable Predictive Challenge. Journal of phycology, 46(2), 220-235. https://doi-org.proxy.lib.uwaterloo.ca/10.1111/j.1529-8817.2010.00815.x

Hammond, J. K., & White, P. R. (1996). The analysis of Non-Stationary Signals Using Time-Frequency Methods. Journal of Sound and Vibration, 190(3), 419-447. https://doi.org/10.1006/jsvi.1996.0072

Hampton, S. E., Moore, M. V., Ozersky, T., Stanley, E. H., Polashenski, C. M., & Galloway, A. W. E. (2015). Heating Up a Cold Subject: Prospects for Under-Ice Plankton Research in Lakes. Journal of Plankton Research, 37(2), 277–284. https://doi.org/10.1093/plankt/fbv002

Han, L. & Jordan, K. J. (2005). Estimating and Mapping Chlorophyll-a Concentration in Pensacola Bay, Florida using Landsat ETM+ data. International Journal of Remote Sensing, 26(23), 5245–5254. https://doi.org/10.1080/01431160500219182

Han, L., Rundquist, D. C., Liu, L. L., Fraser, R. N., & Schalles, J. F. (1994). The Spectral Responses of Algal Chlorophyll in Water with Varying Levels of Suspended Sediment. International Journal of Remote Sensing, 15(18), 3707–3718. https://doi.org/10.1080/01431169408954353

Han, X., Zheng, W., & Wu, C. (2010). Chlorophyll-A Estimation Using Satellite Observations in Tai Lake, China. 2010 International Conference on Multimedia Technology, ICMT 2010, 2, 1–4. https://doi.org/10.1109/ICMULT.2010.5631178

Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., & Medina-Elizade, M. (2006). Global temperature change. Proceedings of the National Academy of Sciences of the United States of America, 103(39), 14288–14293. https://doi.org/10.1073/pnas.0606291103

Harris, T. D., & Smith, V. H. (2016). Do Persistent Organic Pollutants Stimulate Cyanobacterial Blooms?. Inland Waters, 6(2), 124-130. https://doi.org/10.5268/IW-6.2.887

Harrow-Lyle, T. J., & Kirkwood, A. E. (2021). Low Benthic Oxygen and High Internal Phosphorus-Loading are Strongly Associated with the Invasive Macrophyte Nitellopsis obtusa (starry stonewort) in a Large, Polymictic Lake. Frontiers in Environmental Science, 357. https://doi.org/10.3389/fenvs.2021.735509

Hassall, A.H. (1850). A Microscopic Examination of the Water Supplied to the Inhabitants of London and the Suburban Districts. London.

Havens, K., Paerl, H., Phlips, E., Zhu, M., Beaver, J., & Srifa, A. (2016). Extreme Weather Events and Climate Variability Provide a Lens to how Shallow Lakes May Respond to Climate Change. Water, 8(6), 229. https://doi.org/10.3390/w8060229

Hébert, M., Beisner, B. E., Rautio, M., & Fussmann, G. F. (2021). Warming Winters in Lakes: Later Ice Onset Promotes Consumer Overwintering and Shapes Springtime Planktonic Food Webs. Proceedings of the National Academy of Sciences - PNAS, 118(48), 1–. https://doi.org/10.1073/pnas.2114840118

Hennon, G. M. M., & Dyhrman, S. T. (2020). Progress And Promise Of Omics For Predicting The Impacts Of Climate Change On Harmful Algal Blooms. Harmful Algae, 91, 101587–101587. https://doi.org/10.1016/j.hal.2019.03.005

Hensen, V. (1887). Kapitel 1: Über die Bestimmung des Plankton's oder des im Meere treibenden Materials an Pflanzen und Thieren. Jahresbericht der Commission zur Wissenschaftlichen Untersuchung der Deutschen Meere in Kiel: für die Jahre..., 12, 1-107.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X.,

Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 Global Reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hewitt, B. A., Lopez, L. S., Gaibisels, K. M., Murdoch, A., Higgins, S. N., Magnuson, J. J., … Sharma, S. (2018). Historical Trends, Drivers, and Future Projections of Ice Phenoliny in Small North Temperain Lakes in The Laurentian Great Lakes Region. Water (Switzerland), 10(1). https://doi.org/10.3390/w10010070

Hilborn E. D., & Beasley, V. R. (2015). One Health and Cyanobacteria in Freshwater Systems: Animal Illnesses and Deaths Are Sentinel Events for Human Health Risks. Toxins, 7(4), 1374–1395. https://doi.org/10.3390/toxins7041374

Hjerne, O., Hajdu, S., Larsson, U., Downing, A., & Winder, M. (2019). Climate Driven Changes In Timing, Composition And Size Of The Baltic Sea Phytoplankton Spring Bloom. Frontiers in Marine Science, 6(JUL), 1–15. https://doi.org/10.3389/fmars.2019.00482

Ho, J. C., & Michalak, A. M. (2020). Exploring Temperature and Precipitation Impacts On Harmful Algal Blooms Across Continental U.S. Lakes. Limnology and Oceanography, 65(5), 992–1009. https://doi.org/10.1002/lno.11365

Hoekstra, R., Kiers, H. A., & Johnson, A. (2012). Are Assumptions of Well-Known Statistical Techniques Checked, And Why Not? Frontiers in psychology, 3, 137. https://doi.org/10.3389/fpsyg.2012.00137

Holgerson, M. A., Richardson, D. C., Roith, J., Bortolotti, L. E., Finlay, K., Hornbach, D. J., Gurung, K., Ness, A., Andersen, M. R., Bansal, S., Finlay, J. C., Cianci-Gaskill, J. A., Hahn, S., Janke, B. D., McDonald, C., Mesman, J. P., North, R. L., Roberts, C. O., Sweetman, J. N., & Webb, J. R. (2022). Classifying Mixing Regimes in Ponds and Shallow Lakes. Water Resources Research, 58(7). https://doi.org/10.1029/2022WR032522

Hong, H., Shen, R., Zhang, F., Wen, Z., Chang, S., Lin, W., ... & Shi, D. (2017). The Complex Effects of Ocean Acidification On The Prominent N2-Fixing Cyanobacterium Trichodesmium. Science, 356(6337), 527-531. https://doi.org/10.1126/science.aal2981

Hou, Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., & Zheng, C. (2022). Global Mapping Reveals Increase in Lacustrine Algal Blooms Over the Past Decade. Nature Geoscience, 15(2), 130–134. https://doi.org/10.1038/s41561-021-00887-x

Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., & Zheng, C. (2022). Global Mapping Reveals Increase in Lacustrine Algal Blooms Over the Past Decade. Nature Geoscience, 15(2), 130–134. https://doi.org/10.1038/s41561-021-00887-x

Houze, R. A., Wang, J., Fan, J., Brodzik, S., & Feng, Z. (2019). Extreme Convective Storms Over High-Latitude Continental Areas Where Maximum Warming Is Occurring. Geophysical Research Letters, 46(7), 4059–4065. https://doi.org/10.1029/2019GL082414

Howell, S. E. L., Brown, L. C., Kang, K.-K., & Duguay, C. R. (2009). Variability in ice phenology on Great Bear Lake and Great Slave Lake, Northwest Territories, Canada, from SeaWinds/QuikSCAT: 2000–2006. Remote Sensing of Environment, 113(4), 816–834. https://doi.org/10.1016/j.rse.2008.12.007

Hrycik, A. R., Isles, P. D. F., Adrian, R., Albright, M., Bacon, L. C., Berger, S. A., Bhattacharya, R., Grossart, H., Hejzlar, J., Hetherington, A. L., Knoll, L. B., Laas, A., McDonald, C. P., Merrell, K., Nejstgaard, J. C., Nelson, K., Nõges, P., Paterson, A. M., Pilla, R. M., … Pierson, D. C. (2021). Earlier Winter/Spring Runoff and Snowmelt During Warmer Winters Lead To Lower Summer Chlorophyll-A In North Temperate Lakes. Global Change Biology, 27(19), 4615–4629. https://doi.org/10.1111/gcb.15797

Hu, J., Wang, J., Zhang, Z., & Liu, H. (2022). Continuous-Discrete Hybrid Bayesian Network Models for Predicting Earthquake-Induced Liquefaction Based on The Vs Database. Computers & Geosciences, 169, 105231. https://doi.org/10.1016/j.cageo.2022.105231

Huang, J., Zhang, G., Zhang, Y., Guan, X., Wei, Y., & Guo, R. (2020). Global Desertification Vulnerability to Climate Change and Human Activities. Land Degradation and Development, 1–12. https://doi.org/10.1002/ldr.3556

Huang, L., Fang, H., He, G., Jiang, H., & Wang, C. (2016). Effects Of Internal Loading on Phosphorus Distribution in The Taihu Lake Driven by Wind Waves and Lake Currents. Environmental Pollution, 219, 760–773. https://doi.org/10.1016/j.envpol.2016.07.049

Huang, L., Timmermann, A., Lee, S.-S., Rodgers, K. B., Yamaguchi, R., & Chung, E.-S. (2022). Emerging Unprecedented Lake Ice Loss in Climate Change Projections. Nature Communications, 13(1), 5798–5798. https://doi.org/10.1038/s41467-022-33495-3

Huang, S., Li, J., Ye, J., Fleisher, A., Chen, K., Wu, T., & Reiman, E. (2011). Brain Effective Connectivity Modeling for Alzheimer's Disease by Sparse Gaussian Bayesian Network. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 931-939). https://doi.org/10.1145/2020408.2020562

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M., & Visser, P. M. (2018). Cyanobacterial Blooms. Nature Reviews Microbiology, 16(8), 471-483. https://doi.org/10.1038/s41579-018-0040-1

Huisman, J., Matthijs, H. C. P., & Visser, P. M. (2005). Harmful Cyanobacteria. Springer. Dordrecht, Netherlands.

Huo, J., Shi, X., Zhao, S., Sun, B., Li, G., Yu, H., & Wang, S. (2023). Dynamics Of Water Temperature and Interfacial Heat Flux in Lake Ice and Under-Ice Layers. *Hydrology Research*, *54*(7), 855-868. https://doi.org/10.2166/nh.2023.013

Huo, A., Zhang, J., Qiao, C., Li, C., Xie, J., Wang, J., & Zhang, X. (2014). Multispectral Remote Sensing Inversion for City Landscape Water Eutrophication Based on Genetic Algorithm-Support Vector Machine. Water Quality Research Journal of Canada, 49(3), 285–293. https://doi.org/10.2166/wqrjc.2014.040

Hurtado, S. I. RobustLinearReg: Robust Linear Regressions, 2020. URL https://CRAN.R-project.org/package=RobustLinearReg. R package version, 1.2(0), 38.

Ickstadt, K., Bornkamp, B., Grzegorczyk, M., Wieczorek, J., Sheriff, M. R., & Grecco, H. E. (2011). Nonparametric Bayesian Networks (with Discussion) In: Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM, West M, editors. Bayesian Statistics.

Jackson-Blake, L. A., Clayer, F., Haande, S., Sample, J. E., & Moe, S. J. (2022). Seasonal Forecasting of Lake Water Quality And Algal Bloom Risk Using A Continuous Gaussian Bayesian Network. Hydrology and Earth System Sciences, 26(12), 3103–3124. https://doi.org/10.5194/hess-26-3103-2022

Jacobsen, B. A., & Simonsen, P. (1993). Disturbance Events Affecting Phytoplankton Biomass, Composition and Species Diversity in a Shallow, Eutrophic, Temperate Lake. Hydrobiologia, 249(1–3), 9–14. https://doi.org/10.1007/BF00008838

James, R. T., Chimney, M. J., Sharfstein, B., Engstrom, D. R., Schottler, S. P., East, T., & Jin, K. R. (2008). Hurricane Effects on a Shallow Lake Ecosystem, Lake Okeechobee, Florida (USA). Fundamental and Applied Limnology, 172(4), 273–287. https://doi.org/10.1127/1863-9135/2008/0172-0273

Jane, S. F., Mincer, J. L., Lau, M. P., Lewis, A. S. L., Stetler, J. T., & Rose, K. C. (2023). Longer Duration of Seasonal Stratification Contributes to Widespread Increases in Lake Hypoxia and Anoxia. Global Change Biology, 29(4), 1009–1023. https://doi.org/10.1111/gcb.16525

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. Current Opinion in Environmental Sustainability, 36, 1–10. https://doi.org/10.1016/j.cosust.2018.09.001

Jensen, J.R. (2005): Introductory Digital Image Processing: A Remote Sensing Perspective, 4th edn. Prentice-Hall, Englewood Cliffs.

Jeong, J.-H., Kug, J.-S., Linderholm, H. W., Chen, D., Kim, B.-M., & Jun, S.-Y. (2014). Intensified Arctic Warming Under Greenhouse Warming by Vegetation-Atmosphere-Sea Ice Interaction. Environmental Research Letters, 9(9), 94007–. https://doi.org/10.1088/1748-9326/9/9/094007

Jeppesen, E., Kronvang, B., Olesen, J. E., Audet, J., Søndergaard, M., Hoffmann, C. C., Andersen, H. E., Lauridsen, T. L., Liboriussen, L., Larsen, S. E., Beklioglu, M., Meerhoff, M., Özen, A., & Özkan, K. (2011). Climate Change Effects on Nitrogen Loading from Cultivated Catchments In Europe: Implications For Nitrogen Retention, Ecological State Of Lakes And Adaptation. Hydrobiologia, 663(1), 1–21. https://doi.org/10.1007/s10750-010-0547-6

Jewson, D. H., Granin, N. G., Zhdanov, A. A., & Gnatovsky, R. Y. (2009). Effect of Snow Depth on Under-Ice Irradiance and Growth of Aulacoseira Baicalensis In Lake Baikal. Aquatic Ecology, 43(3), 673–679. https://doi.org/10.1007/s10452-009-9267-2

Jiang, D., Scholze, J., Liu, X., Simis, S. G., Stelzer, K., Müller, D., ... & Spyrakos, E. (2023). A Data-Driven Approach to Flag Land-Affected Signals in Satellite Derived Water Quality from Small Lakes. International Journal of Applied Earth Observation and Geoinformation, 117, 103188. https://doi.org/10.1016/j.jag.2023.103188

Jiang, P., Liu, X., Zhang, J., Te, S. H., Gin, K. Y. H., Van Fan, Y., ... & Shoemaker, C. A. (2021). Cyanobacterial Risk Prevention Under Global Warming Using an Extended Bayesian Network. Journal Of Cleaner Production, 312, 127729. https://doi.org/10.1016/j.jclepro.2021.127729

Jirsa, F., Gruber, M., Stojanovic, A., Omondi, S. O., Mader, D., Körner, W., & Schagerl, M. (2013). Major And Trace Element Geochemistry of Lake Bogoria and Lake Nakuru, Kenya, During Extreme Draught. Chemie Der Erde, 73(3), 275–282. https://doi.org/10.1016/j.chemer.2012.09.001

Johansen, R. A., Katzenmeyer, A. W., Pokrzywinski, K. L., & Reif, M. K. (2023). A Review of Sensor-Based Approaches for Monitoring Rapid Response Treatments of cyanoHABs.

Johnson, L. (1975). Distribution Of Fish Species In Great Bear Lake, Northwest Territories, With Reference To Zooplankton, Benthic Invertebrates, And Environmental Conditions. Journal of the Fisheries Board of Canada, 32(11), 1989-2004. https://doi.org/10.1139/f75-235

Jones, G. J., & Poplawski, W. (1998). Understanding And Management of Cyanobacterial Blooms In Sub-Tropical Reservoirs Of Queensland, Australia. Water Science and Technology, 37(2), 161-168.

Joung, D., Leduc, M., Ramcharitar, B., Xu, Y., Isles, P. D., Stockwell, J. D., ... & Schroth, A. W. (2017). Winter Weather and Lake-Watershed Physical Configuration Drive Phosphorus, Iron, And Manganese Dynamics In Water And Sediment Of Ice-Covered Lakes. Limnology and Oceanography, 62(4), 1620-1635. https://doi.org/10.1002/lno.10521

Kalinowska, K., Napiórkowska-Krzebietke, A., Bogacka-Kapusta, E., & Stawecki, K. (2019). Comparison Of Ice-On and Ice-Off Abiotic and Biotic Parameters in Three

Eutrophic Lakes. Ecological Research, 34(5), 687–698. https://doi.org/10.1111/1440-1703.12039

Kämpf. J. (2017). Wind-Driven Overturning, Mixing and Upwelling in Shallow Water: A Nonhydrostatic Modeling Study. Journal of Marine Science and Engineering, 5(4), 47–. https://doi.org/10.3390/jmse5040047

Kane, D. D., Conroy, J. D., Richards, R. P., Baker, D. B., & Culver, D. A. (2014). Re-Eutrophication of Lake Erie: Correlations Between Tributary Nutrient Loads and Phytoplankton Biomass. Journal of Great Lakes Research, 40(3), 496-501. https://doi.org/10.1016/j.jglr.2014.04.004

Karatayev, A. Y., Burlakova, L. E., Mehler, K., Hinchey, E. K., Wick, M., Bakowska, M., & Mrozinska, N. (2021). Rapid Assessment of Dreissena Population In Lake Erie Using Underwater Videography. Hydrobiologia, 848(9), 2421-2436. https://doi.org/10.1007/s10750-020-04481-x

Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., & Senior, C. A. (2014). Heavier Summer Downpours with Climate Change Revealed by Weather Forecast Resolution Model. Nature Climate Change, 4(7), 570–576. https://doi.org/10.1038/nclimate2258

Kendrick, M. R., Hershey, A. E., & Huryn, A. D. (2019). Disturbance, Nutrients, and Antecedent Flow Conditions Affect Macroinvertebrate Community Structure and Productivity in an Arctic River. Limnology and Oceanography, 64(S1), S93-S104. https://doi.org/10.1002/lno.10942

Khakzad, N., Khan, F., & Amyotte, P. (2013). Risk-Based Design of Process Systems Using Discrete-Time Bayesian Networks. Reliability Engineering & System Safety, 109, 5-17. https://doi.org/10.1016/j.ress.2012.07.009

Kim, T.-J. (2018). Prevention of Harmful Algal Blooms by Control of Growth Parameters. Advances in Bioscience and Biotechnology, 09(11), 613–648. https://doi.org/10.4236/abb.2018.911043

Kim, Y. H., Im, J., Ha, H. K., Choi, J.-K., & Ha, S. (2014). Machine Learning Approaches to Coastal Water Quality Monitoring Using GOCI Satellite Data. GIScience and Remote Sensing, 51(2), 158–174. https://doi.org/10.1080/15481603.2014.900983

Kimmel, D. G., Boicourt, W. C., Pierson, J. J., Roman, M. R., & Zhang, X. (2009). A Comparison of The Mesozooplankton Response To Hypoxia In Chesapeake Bay And The Northern Gulf Of Mexico Using The Biomass Size Spectrum. Journal of Experimental Marine Biology and Ecology, 381(SUPPL.), S65–S73. https://doi.org/10.1016/j.jembe.2009.07.012

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arxiv.1412.6980

Klais, R., Tamminen, T., Kremp, A., Spilling, K., & Olli, K. (2011). Decadal-scale changes of Dinoflagellates and Diatoms in the Anomalous Baltic Sea spring bloom. PLoS ONE, 6(6). https://doi.org/10.1371/journal.pone.0021567

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-Normalizing Neural Networks. Advances In Neural Information Processing Systems, 30. https://doi.org/10.48550/arxiv.1706.02515

Koller, D., & Friedman, N. (2009). Probabilistic Graphical Models: Principles And Techniques. MIT press

Komatsu, E., Fukushima, T., & Harasawa, H. (2007). A Modeling Approach To Forecast The Effect Of Long-Term Climate Change On Lake Water Quality. Ecological Modelling, 209(2), 351–366. https://doi.org/10.1016/j.ecolmodel.2007.07.021

Kong, J. L., Sun, X. M., Wong, D. W., Chen, Y., Yang, J., Yan, Y., & Wang, L. X. (2015). A Semi-Analytical Model for Remote Sensing Retrieval of Suspended Sediment Concentration in the Gulf of Bohai, China. Remote Sensing, 7(5), 5373–5397. https://doi.org/10.3390/rs70505373

Kong, X., He, Q., Yang, B., He, W., Xu, F., Janssen, A. B. G. … Mooij, W. M. (2017a). Hydrological Regulation Drives Regime Shifts: Evidence from Paleolimnology and Ecosystem Modeling of a Large Shallow Chinese Lake. Global Change Biology, 23(2), 737–754. https://doi.org/10.1111/gcb.13416

Kong, X., Sun, Y., Su, R., & Shi, X. (2017b). Real-Time Eutrophication Status Evaluation of Coastal Waters Using Support Vector Machine with Grid Search Algorithm. Marine Pollution Bulletin, 119(1), 307–319. https://doi.org/10.1016/j.marpolbul.2017.04.022

Konik, M., Kowalewski, M., Bradtke, K., & Darecki, M. (2019). the Operational Method of Filling Information Gaps in Satellite Imagery Using Numerical Models. International Journal of Applied Earth Observation and Geoinformation, 75, 68-82. https://doi.org/10.1016/j.jag.2018.09.002

Koushik, A. N., Manoj, M., & Nezamuddin, N. (2020). Machine Learning Applications in Activity-Travel Behaviour Research: a Review. Transport Reviews, 40(3), 288-311.3 https://doi.org/10.1080/01441647.2019.1704307

Kraemer, B. M., Anneville, O., Chandra, S., Dix, M., Kuusisto, E., Livingstone, D. M., … Mcintyre, P. B. (2015). Morphometry and Average Temperature Affect Lake Stratification Responses to Climate Change. Geophysical Research Letters, 42(12), 4981–4988. https://doi.org/10.1002/2015GL064097

Kraemer, B. M., Kakouei, K., Munteanu, C., Thayne, M. W., & Adrian, R. (2022). Worldwide Moderate-Resolution Mapping of Lake Surface Chl-A Reveals Variable Responses to Global Change (1997–2020). PLOS Water, 1(10), e0000051. https://doi.org/10.1371/journal.pwat.0000051

Kraemer, Mehner, T., & Adrian, R. (2017). Reconciling the Opposing Effects of Warming on Phytoplankton Biomass in 188 Large Lakes. Scientific Reports, 7(1), 10762–10767. https://doi.org/10.1038/s41598-017-11167-3

Kulshreshtha, S. (2020). An ARIMA-LSTM Hybrid Model for Stock Market Prediction Using Live Data. Journal of Engineering Science & Technology Review, 13(4). https://doi.org/10.3390/rs9050420

Kurek, M. R., Frey, K. E., Guillemette, F., Podgorski, D. C., Townsend-Small, A., Arp, C. D., Kellerman, A. M., & Spencer, R. G. M. (2022). Trapped Under Ice: Spatial and Seasonal Dynamics of Dissolved Organic Matter Composition in Tundra Lakes. Journal of Geophysical Research. Biogeosciences, 127(4). https://doi.org/10.1029/2021JG006578

Kwisthout, J. (2018). Approximate inference in Bayesian networks: Parameterized Complexity Results. International Journal of Approximate Reasoning, 93, 119-131. https://doi.org/10.1016/j.ijar.2017.10.029

Kyrimi, E., McLachlan, S., Dube, K., Neves, M. R., Fahmi, A., & Fenton, N. (2021). a Comprehensive Scoping Review of Bayesian Networks in Healthcare: Past, Present and Future. Artificial Intelligence in Medicine, 117, 102108. https://doi.org/10.1016/j.artmed.2021.102108

Lackey, & Sawyer, C. N. (1945). Plankton Productivity of Certain South-Eastern Wisconsin Lakes as Related to Fertilization: I. Surveys. Sewage Works Journal, 17(3), 573–585.

Lahive, E., O' Halloran, J., & Jansen, M. A. K. (2011). Differential Sensitivity of Four Lemnaceae Species to Zinc Sulphate. Environmental and Experimental Botany, 71(1), 25–33. https://doi.org/10.1016/j.envexpbot.2010.10.014

Lake Winnipeg Implementation Committee (LWIC) (2005). Restoring the Health of Lake Winnipeg-A Report by the Lake Winnipeg Implementation Committee. Government of Manitoba, Government of Canada.

Lamb, M. C., Kimmel, D. G., & Field, E. K. (2019). The Effects of Temperature on Bosmina Longirostris Susceptibility to Microcystin-LR Acute Toxicity. PLoS ONE, 14(7), 1–15. https://doi.org/10.1371/journal.pone.0219342

Lathrop, R. G., Lillesand, T. M., & Yandell, B. S. (1991). Testing the Utility of Simple Multi-Date Thematic Mapper Calibration Algorithms for Monitoring Turbid Inland Waters. International Journal of Remote Sensing, 12(10), 2045–2063. https://doi.org/10.1080/01431169108955235

Law, J., Quick, M., & Chan, P. W. (2014). Analyzing Hotspots of Crime Using a Bayesian Spatiotemporal Modeling Approach: a Case Study of Violent Crime in the Greater Toronto Area. Geographical Analysis, 47(1), 1–19. https://doi.org/10.1111/gean.12047

Leconte, R., Temimi, M., Chaouch, N., Brissette, F., & Toussaint, T. (2008). on the Use of Satellite Passive Microwave Data for Estimating Surface Soil Wetness in the

Mackenzie River Basin. in Cold Region Atmospheric and Hydrologic Studies. the Mackenzie GEWEX Experience (pp. 59-79). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-75136-6_4

Leggesse, E. S., Zimale, F. A., Sultan, D., Enku, T., Srinivasan, R., & Tilahun, S. A. (2023). Predicting Optical Water Quality Indicators from Remote Sensing Using Machine Learning Algorithms in Tropical Highlands of Ethiopia. Hydrology, 10(5), 110. https://doi.org/10.3390/hydrology10050110

Lehman, P. W., Kurobe, T., Lesmeister, S., Baxa, D., Tung, A., & Teh, S. J. (2017). Impacts of the 2014 Severe Drought on the Microcystis Bloom in San Francisco Estuary. Harmful Algae, 63, 94–108. https://doi.org/10.1016/j.hal.2017.01.011

Lehner, B., Verdin, K., Jarvis, A. (2008). New Global Hydrography Derived from Spaceborne Elevation Data. Eos, Transactions, American Geophysical Union, 89(10): 93–94. https://doi.org/10.1029/2008eo100001

Lerner, U., Segal, E., & Koller, D. (2013). Exact Inference in Networks with Discrete Children of Continuous Parents. arXiv preprint arXiv:1301.2289. https://doi.org/10.48550/arxiv.1301.2289

Lévesque, L. M. J., Roy, J., Glozier, N. E., Dirk, L., & Cooke, C. A. (2023). Dissolved polycyclic aromatic compounds in Canada's Athabasca River in relation to Oil Sands from 2013 through 2019. *Environmental Monitoring and Assessment*, *195*(11), 1354–1354. https://doi.org/10.1007/s10661-023-11846-x

Li, J., Gao, M., Feng, L., Zhao, H., Shen, Q., Zhang, F., ... & Zhang, B. (2019). Estimation of chlorophyll-a Concentrations in a Highly Turbid Eutrophic Lake Using a Classification-Based MODIS Land-Band Algorithm. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(10), 3769-3783.

Li, W., Xu, X., Fujibayashi, M., Niu, Q., Tanaka, N., & Nishimura, O. (2016). Response of Microalgae to Elevated CO2 and Temperature: Impact of Climate Change on Freshwater Ecosystems. Environmental Science and Pollution Research, 23(19), 19847–19860. https://doi.org/10.1007/s11356-016-7180-5

Li, X., Peng, S., Xi, Y., Woolway, R. I., & Liu, G. (2022). Earlier Ice Loss Accelerates Lake Warming in the Northern Hemisphere. Nature Communications, 13(1), 5156. https://doi.org/10.1038/s41467-022-32830-y

Li, Y., Lu, W., & Wang, H. (2007). Inversing chlorophyll-a Concentration by Multi-Temporal Models Using TM Images. 27th International Congress on High-Speed Photography and Photonics, 6279(January 2007), 62795H. https://doi.org/10.1117/12.725430

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

Lin, S., Litaker, R. W., & Sunda, W. G. (2016). Phosphorus Physiological Ecology and Molecular Mechanisms in Marine Phytoplankton. Journal of Phycology, 52(1), 10–36. https://doi.org/10.1111/jpy.12365

Lin, S., Pierson, D. C., & Mesman, J. P. (2023). Prediction of Algal Blooms Via Data-Driven Machine Learning Models: an Evaluation Using Data from a Well-Monitored Mesotrophic Lake. Geoscientific Model Development, 16(1), 35-46. https://doi.org/10.5194/gmd-16-35-2023

Litke, D. W. (1999). Review of Phosphorus Control Measures in the United States and Their Effects on Water Quality (Vol. 99, No. 4007). US Department of the Interior, US Geological Survey.

Liu, G., Li, Y., Lyu, H., Wang, S., Du, C., & Huang, C. (2016). an Improved Land Target-Based Atmospheric Correction Method for Lake Taihu. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(2), 793–803. https://doi.org/10.1109/JSTARS.2015.2503800

Liu, J., Kang, Y., Liu, K., Yang, X., Sun, M., & Hu, J. (2020). Maize Carotenoid Gene Locus Mining Based on Conditional Gaussian Bayesian Network. IEEE Access, 8, 15223-15231. https://doi.org/10.1109/ACCESS.2020.2966590

Liu, S., & Duan, A. (2018). Impacts of the Global Sea Surface Temperature Anomaly on the Evolution of Circulation and Precipitation in East Asia on a Quasi-Quadrennial Cycle. Climate Dynamics, 51(11-12), 4077–4094. https://doi.org/10.1007/s00382-017-3663-4

Liu, W., Shao, Y., Li, C., Li, C., & Jiang, Z. (2023). Development of a Non-Gaussian Copula Bayesian Network for Safety Assessment of Metro Tunnel Maintenance. Reliability Engineering & System Safety, 109423. https://doi.org/10.1016/j.ress.2023.109423

Lopez, L. S., Hewitt, B. A., & Sharma, S. (2019). Reaching a Breaking Point: How is Climate Change Influencing the Timing of Ice Breakup in Lakes Across the Northern Hemisphere? Limnology and Oceanography, 2621–2631. https://doi.org/10.1002/lno.11239

Ludsin, S. A., Kershner, M. W., Blocksom, K. A., Knight, R. L., & Stein, R. A. (2001). Life After Death in Lake Erie: Nutrient Controls Drive Fish Species Richness, Rehabilitation. Ecological Applications, 11(3), 731-746.

Lui, Li, W. K., Leung, K. M. Y., Lee, J. H. W., & Jayawardena, A. W. (2007). Modelling Algal Blooms Using Vector Autoregressive Model with Exogenous Variables and Long Memory Filter. Ecological Modelling, 200(1), 130–138. https://doi.org/10.1016/j.ecolmodel.2006.06.017

Luo, Y. W., Shi, D., Kranz, S. A., Hopkinson, B. M., Hong, H., Shen, R., & Zhang, F. (2019). Reduced Nitrogenase Efficiency Dominates Response of the Globally Important Nitrogen Fixer Trichodesmium to Ocean Acidification. Nature Communications, 10(1), 1–12. https://doi.org/10.1038/s41467-019-09554-7

Lürling, M., Van Oosterhout, F., & Faassen, E. (2017). Eutrophication and Warming Boost Cyanobacterial Biomass and Microcystins. Toxins, 9(2), 1–16. https://doi.org/10.3390/toxins9020064

Lytvynenko, V., Nikytenko, D., Voronenko, M., Savina, N., & Naumov, O. (2020). Assessing the Possibility of a Country's Economic Growth Using Dynamic Bayesian Network Models. in 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 1, pp. 36-39). IEEE. https://doi.org/10.1109/CSIT49958.2020.9321995

Maberly, S. C., O'Donnell, R. A., Woolway, R. I., Cutler, M. E. J., Gong, M., Jones, I. D., … Tyler, A. N. (2020). Global Lake Thermal Regions Shift Under Climate Change. Nature Communications, 11(1), 1–9. https://doi.org/10.1038/s41467-020-15108-z

Maki, A. W., Porcella, D. B., & Wendt, R. H. (1984). The Impact of Detergent Phosphorus Bans on Receiving Water Quality. Water Research, 18(7), 893–903. https://doi.org/10.1016/0043-1354(84)90275-6

Maltsev, Y., Maltseva, K., Kulikovskiy, M., & Maltseva, S. (2021). Influence of Light Conditions on Microalgae Growth and Content of Lipids, Carotenoids, and Fatty Acid Composition. Biology (Basel, Switzerland), 10(10), 1060–. https://doi.org/10.3390/biology10101060

Mani, N., & Srinivasan, B. (1997). Application of Artificial Neural Network Model for Optical Character Recognition. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 3, 2517–2520. https://doi.org/10.1109/icsmc.1997.635312

Manivanan, R. (2008). Water Quality Modeling: Rivers, Streams, and Estuaries. New India Publishing.

Marella, D., Vicard, P., Vitale, V., & Ababei, D. (2019). Measurement Error Correction by Nonparametric Bayesian Networks: Application and Evaluation. in Statistical Learning of Complex Data (pp. 155–162). Springer International Publishing. https://doi.org/10.1007/978-3-030-21140-0_16

Masmoudi, K., Abid, L., & Masmoudi, A. (2019). Credit Risk Modeling Using Bayesian Network with a Latent Variable. Expert Systems with Applications, 127, 157-166. https://doi.org/10.1016/j.eswa.2019.03.014

Mathews, A., Singh, K. K., Cummings, A. R., & Rogers, S. R. (2023). Fundamental Practices for Drone Remote Sensing Research Across Disciplines. Drone Systems and Applications. https://doi.org/10.1139/dsa-2023-0021

Matthews, M. W. (2011). A Current Review of Empirical Procedures of Remote Sensing in Inland and Near-Coastal Transitional Waters. International Journal of Remote Sensing, 32(21), 6855–6899. https://doi.org/10.1080/01431161.2010.512947

Matthews, M. W., & Odermatt, D. (2015). Improved Algorithm for Routine Monitoring of Cyanobacteria and Eutrophication in Inland and Near-Coastal Waters. Remote Sensing of Environment, 156, 374–382. https://doi.org/10.1016/j.rse.2014.10.010

Matthews, M. W., Bernard, S., & Robertson, L. (2012). An Algorithm for Detecting Trophic Status (Chlorophyll-A), Cyanobacterial-Dominance, Surface Scums and Floating Vegetation in Inland and Coastal Waters. Remote Sensing of Environment, 124, 637–652. https://doi.org/10.1016/j.rse.2012.05.032

Mayer, B., & Kylling, A. (2005). Technical note: the Libradtran Software Package for Radiative Transfer Calculations - Description and Examples of Use. Atmospheric Chemistry and Physics, 5(7), 1855–1877. https://doi.org/10.5194/acp-5-1855-2005

McCrystall, M. R., Stroeve, J., Serreze, M., Forbes, B. C., & Screen, J. A. (2021). New Climate Models Reveal Faster and Larger Increases in Arctic Precipitation Than Previously Projected. Nature Communications, 12(1), 6765–12. https://doi.org/10.1038/s41467-021-27031-y

McElroy, T. (2012). An Alternative Model-Based Seasonal Adjustment That Reduces Residual Seasonal Autocorrelation. Taiwan Economic Forecast and Policy, 43(1), 33.

McLachlan, S., Dube, K., Hitman, G. A., Fenton, N. E., & Kyrimi, E. (2020). Bayesian Networks in Healthcare: Distribution by Medical Condition. Artificial Intelligence in Medicine, 107, 101912. Https://Doi.Org/10.1016/J.Artmed.2020.101912

McLeod, A. I. (2005). Kendall Rank Correlation and Mann-Kendall Trend Test. R package Kendall, 602, 1-10.

Meineri, E., Dahlberg, C. J., & Hylander, K. (2015). Using Gaussian Bayesian Networks to Disentangle Direct and Indirect Associations Between Landscape Physiography, Environmental Variables and Species Distribution. Ecological Modelling, 313, 127-136. https://doi.org/10.1016/j.ecolmodel.2015.06.028

Mesman, J. P., Stelzer, J. A. A., Dakos, V., Goyette, S., Jones, I. D., Kasparian, J., McGinnis, D. F., & Ibelings, B. W. (2021). the Role of Internal Feedbacks in Shifting Deep Lake Mixing Regimes Under a Warming Climate. Freshwater Biology, 66(6), 1021–1035. https://doi.org/10.1111/fwb.13704

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., ... & Zagorski, M. A. (2013). Record-Setting Algal Bloom in Lake Erie Caused by Agricultural and Meteorological Trends Consistent with Expected Future Conditions. Proceedings of the National Academy of Sciences, 110(16), 6448-6452. https://doi.org/10.1073/pnas.1216006110

Mika, A. M. (1997). Three decades of Landsat instruments. Photogrammetric Engineering and Remote Sensing, 63(7), 839-852.

Miller, M. P., McKnight, D. M., & Chapra, S. C. (2009). Production of Microbially-Derived Fulvic Acid from Photolysis of Quinone-Containing Extracellular Products of

Phytoplankton. Aquatic Sciences, 71(2), 170–178. https://doi.org/10.1007/s00027-009-9194-2

Min, S. K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human Contribution to More-Intense Precipitation Extremes. Nature, 470(7334), 378–381. https://doi.org/10.1038/nature09763

Mitchell, & Prepas, E. E. (1990). the Atlas of Alberta Lakes. University of Alberta Press.

Mitsch, W. J. (2017). Solving Lake Erie's Harmful Algal Blooms by Restoring the Great Black Swamp in Ohio. Ecological engineering, 108, 406-413. https://doi.org/10.1016/j.ecoleng.2017.08.040

Moe, S. J., Haande, S., & Couture, R. M. (2016). Climate Change, Cyanobacteria Blooms and Ecological Status of Lakes: a Bayesian Network Approach. Ecological Modelling, 337, 330-347. https://doi.org/10.1016/j.ecolmodel.2016.07.004

Mohebzadeh, H., Mokari, E., Daggupati, P., & Biswas, A. (2021). A Machine Learning Approach for Spatiotemporal Imputation of MODIS Chlorophyll-A. International Journal of Remote Sensing, 42(19), 7381–7404. https://doi.org/10.1080/01431161.2021.1957513

Molina, J., Pulido-Velázquez, D., García-Aróstegui, J. L., & Pulido-Velázquez, M. (2013). Dynamic Bayesian Networks as a Decision Support Tool for Assessing Climate Change Impacts on Highly Stressed Groundwater Systems. Journal of Hydrology, 479, 113–129. https://doi.org/10.1016/j.jhydrol.2012.11.038

Molkov, A., Fedorov, S., & Pelevin, V. (2022). Toward Atmospheric Correction Algorithms for Sentinel-3/OLCI Images of Productive Waters. Remote Sensing (Basel, Switzerland), 14(15), 3663–. https://doi.org/10.3390/rs14153663

Molot, L. A., Watson, S. B., Creed, I. F., Trick, C. G., Mccabe, S. K., Verschoor, M. J., … Schiff, S. L. (2014). A Novel Model for Cyanobacteria Bloom Formation: the Critical Role of Anoxia and Ferrous Iron. Freshwater Biology, 59(6), 1323–1340. https://doi.org/10.1111/fwb.12334

Moody, C. S., & Worrall, F. (2017). Modeling Rates of DOC Degradation Using DOM Composition and Hydroclimatic Variables. Journal of Geophysical Research: Biogeosciences, 122(5), 1175–1191. https://doi.org/10.1002/2016JG003493

Moore, S. K., Cline, M. R., Blair, K., Klinger, T., Varney, A., & Norman, K. (2019). an Index of Fisheries Closures Due to Harmful Algal Blooms and a Framework for Identifying Vulnerable Fishing Communities on the U.S. West Coast. Marine Policy, 110, 103543. https://doi.org/10.1016/j.marpol.2019.103543

Moore, T. S., Dowell, M. D., Bradt, S., & Ruiz Verdu, A. (2014). an Optical Water Type Framework for Selecting and Blending Retrievals from Bio-Optical Algorithms in Lakes and Coastal Waters. Remote Sensing of Environment, 143, 97–111. https://doi.org/10.1016/j.rse.2013.11.021

Moore, T. S., Mouw, C. B., Sullivan, J. M., Twardowski, M. S., Burtner, A. M., Ciochetto, A. B., … Weidemann, A. (2017). Bio-Optical Properties of Cyanobacteria Blooms in Western Lake Erie. Frontiers in Marine Science, 4(SEP), 1–20. https://doi.org/10.3389/fmars.2017.00300

Morel, A., & Prieur, L. (1977). Analysis of Variations in Ocean Color. Limnology and Oceanography, 22(4), 709–722. https://doi.org/10.4319/lo.1977.22.4.0709

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). the Fallacy of Placing Confidence in Confidence Intervals. Psychonomic Bulletin & Review, 23, 103-123. https://doi.org/10.3758/s13423-015-0947-8

Mosley. (2015). Drought Impacts on the Water Quality of Freshwater Systems; Review and Integration. Earth-Science Reviews, 140, 203–214. https://doi.org/10.1016/j.earscirev.2014.11.010

Moss, B., McKee, D., Atkinson, D., Collings, S. E., Eaton, J. W., Gill, A. B., Harvey, I., Hatton, K., Heyes, T., & Wilson, D. (2003). How Important is Climate? Effects of Warming, Nutrient Addition and Fish on Phytoplankton in Shallow Lake Microcosms. the Journal of Applied Ecology, 40(5), 782–792. https://doi.org/10.1046/j.1365-2664.2003.00839.x

Moss, B., McKee, D., Atkinson, D., Collings, S. E., Eaton, J. W., Gill, A. B., Harvey, I., Hatton, K., Heyes, T., & Wilson, D. (2003). How Important is Climate? Effects of Warming, Nutrient Addition and Fish on Phytoplankton in Shallow Lake Microcosms. the Journal of Applied Ecology, 40(5), 782–792. https://doi.org/10.1046/j.1365-2664.2003.00839.x

Moss, B., Stephen, D., Balayla, D. M., Bécares, E., Collings, S. E., Fernández-Aláez, C., Fernández-Aláez, M., Ferriol, C., García, P., Gomá, J., Gyllström, M., Hansson, L.-A., Hietala, J., Kairesalo, T., Miracle, M. R., Romo, S., Rueda, J., Russell, V., Ståhl-Delbanco, A., … Villena, M. J. (2004). Continental-Scale Patterns of Nutrient and Fish Effects on Shallow Lakes: Synthesis of a Pan-European Mesocosm Experiment. Freshwater Biology, 49(12), 1633–1649. https://doi.org/10.1111/j.1365-2427.2004.01304.x

Muir, A. M., Leonard, D. M., & Krueger, C. C. (2013). Past, Present and Future of Fishery Management on One of the World's Last Remaining Pristine Great Lakes: Great Bear Lake, Northwest Territories, Canada. Reviews in Fish Biology and Fisheries, 23(3), 293–315. https://doi.org/10.1007/s11160-012-9295-1

Müller-Wilm, U. (2017). Sen2Cor Configuration and User Manual. Telespazio VEGA Deutschland GmbH, (1), 53. Retrieved from https://step.esa.int/thirdparties/sen2cor/2.4.0/Sen2Cor_240_Documenation_PDF/S2-PDGS-MPC-L2A-SUM-V2.4.0.pdf

Muñoz Sabater, J., (2019): ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on < 04-21-2022 >), 10.24381/cds.e2161bac

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J. N. (2021). ERA5-Land: a State-Of-The-Art Global Reanalysis Dataset for Land Applications, Earth Syst. Sci. Data,13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.

Mustafa, F., Bu, L., Wang, Q., Yao, N., Shahzaman, M., Bilal, M., Rana, W. A., & Iqbal, R. (2021). Neural-Network-Based Estimation of Regional-Scale Anthropogenic CO2 Emissions Using an Orbiting Carbon Observatory-2 (OCO-2) Dataset Over East and West Asia. Atmospheric Measurement Techniques, 14(11), 7277–7290. https://doi.org/10.5194/amt-14-7277-2021

Myrstener, M., Fork, M. L., Bergström, A. K., Puts, I. C., Hauptmann, D., Isles, P. D., ... & Sponseller, R. A. (2022). Resolving the Drivers of Algal Nutrient Limitation from Boreal to Arctic Lakes and Streams. Ecosystems, 25(8), 1682-1699. https://doi.org/10.1007/s10021-022-00759-4

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in R with applications in systems biology. Springer.

Neapolitan, R. E. (2004). Learning bayesian networks (Vol. 38). Upper Saddle River: Pearson Prentice Hall.

Neil, C., Spyrakos, E., Hunter, P. D., & Tyler, A. N. (2019). a Global Approach for chlorophyll-a Retrieval Across Optically Complex Inland Waters Based on Optical Water Types. Remote Sensing of Environment, 229, 159–178. https://doi.org/10.1016/j.rse.2019.04.027

Nicholson, W. B., Matteson, D. S., & Bien, J. (2017). VARX-L: Structured Regularization for Large Vector Autoregressions with Exogenous Variables. International Journal of Forecasting, 33(3), 627-651. https://doi.org/10.1016/j.ijforecast.2017.01.003

Niedrist, G. H., Psenner, R., & Sommaruga, R. (2018). Climate Warming Increases Vertical and Seasonal Water Temperature Differences and Inter-Annual Variability in a Mountain Lake. Climatic Change, 151(3–4), 473–490. https://doi.org/10.1007/s10584-018-2328-6

Nofrizal, A. Y., Sonobe, R., Yamashita, H., Seki, H., Mihara, H., Morita, A., & Ikka, T. (2022). Evaluation of a One-Dimensional Convolution Neural Network for Chlorophyll Content Estimation Using a Compact Spectrometer. Remote Sensing, 14(9), 1997. https://doi.org/10.3390/rs14091997

Nõges, P., Adrian, R., Anneville, O., Arvola, L., Blenckner, T., George, G., ... & Weyhenmeyer, G. (2010). the Impact of Variations in the Climate on Seasonal Dynamics of Phytoplankton (Pp. 253-274). Springer Netherlands. https://doi.org/10.1007/978-90-481-2945-4_14

Nolan, S., Bollens, S. M., & Rollwagen-Bollens, G. (2019). Diverse Taxa of Zooplankton Inhabit Hypoxic Waters During Both Day and Night in a Temperate Eutrophic Lake. Journal of Plankton Research, 41(4), 431–447. https://doi.org/10.1093/plankt/fbz021

Nürnberg, G. K., & LaZerte, B. D. (2016). More Than 20 Years of Estimated Internal Phosphorus Loading in Polymictic, Eutrophic Lake Winnipeg, Manitoba. Journal of Great Lakes Research, 42(1), 18-27. https://doi.org/10.1016/j.jglr.2015.11.003

O'Neil, J. M., Davis, T. W., Burford, M. A., & Gobler, C. J. (2012). the Rise of Harmful Cyanobacteria Blooms: The Potential Roles of Eutrophication and Climate Change. Harmful Algae, 14, 313–334. https://doi.org/10.1016/j.hal.2011.10.027

O'Reilly, C. M., Sharma, S., Gray, D. K., Hampton, S. E., Read, J. S., Rowley, R. J., Schneider, P., Lenters, J. D., McIntyre, P. B., Kraemer, B. M., Weyhenmeyer, G. A., Straile, D., Dong, B., Adrian, R., Allan, M. G., Anneville, O., Arvola, L., Austin, J., Bailey, J. L., … Dokulil, M. T. (2015). Rapid and Highly Variable Warming of Lake Surface Waters Around the Globe. Geophysical Research Letters, 42(24), 10,773–10,781. https://doi.org/10.1002/2015GL066235

O'Reilly, J. E., & Werdell, P. J. (2019). Chlorophyll-algorithms for Ocean Color Sensors - OC4, OC5 & OC6. Remote Sensing of Environment, 229, 32–47. https://doi.org/10.1016/j.rse.2019.04.021

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., & McClain, C. (1998). Ocean Color Chlorophyll-algorithms for Seawifs. Journal of Geophysical Research, Washington, DC, 103(C11), 24937–24953. https://doi.org/10.1029/98JC02160

Obryk, M. K., Doran, P. T., & Priscu, J. C. (2019). Prediction of Ice-Free Conditions for a Perennially Ice-Covered Antarctic Lake. Journal of Geophysical Research: Earth Surface, 124(2), 686–694. https://doi.org/10.1029/2018JF004756

Odermatt, D., Gitelson, A., Brando, V. E., & Schaepman, M. (2012). Review of Constituent Retrieval in Optically Deep and Complex Waters from Satellite Imagery. Remote sensing of environment, 118, 116-126. https://doi.org/10.1016/j.rse.2011.11.013

Oehmcke, S., Zielinski, O., & Kramer, O. (2016). kNN Ensembles with Penalized DTW for Multivariate Time Series Imputation. 2016 International Joint Conference on Neural Networks (IJCNN), 2016-, 2774–2781. https://doi.org/10.1109/IJCNN.2016.7727549

Ogashawara, I., Mishra, D. R., & Gitelson, A. A. (2017). Remote Sensing of Inland Waters: Background and Current State-Of-The-Art. in Bio-Optical Modeling and Remote Sensing of Inland Waters (pp. 1-24). Elsevier.

Ohki, K. (1976). Effect of Zinc Nutrition on Photosynthesis and Carbonic Anhydrase Activity in Cotton. Physiologia Plantarum, 38(4), 300–304. https://doi.org/10.1111/j.1399-3054.1976.tb04007.x

Orihel, D. M., Baulch, H. M., Casson, N. J., North, R. L., Parsons, C. T., Seckar, D. C. M., & Venkiteswaran, J. J. (2017). Internal Phosphorus Loading in Canadian Fresh Waters: a Critical Review and Data Analysis. Canadian Journal of Fisheries and Aquatic Sciences, 74(12), 2005–2029. https://doi.org/10.1139/cjfas-2016-0500

Orihel, Schindler, D. W., Ballard, N. C., Graham, M. D., O'Connell, D. W., Wilson, L. R., & Vinebrooke, R. D. (2015). the "Nutrient Pump": Iron-Poor Sediments Fuel Low Nitrogen-To-Phosphorus Ratios and Cyanobacterial Blooms in Polymictic Lakes. Limnology and Oceanography, 60(3), 856–871. https://doi.org/10.1002/lno.10076

Paerl, H. W. (2018). Mitigating Toxic Planktonic Cyanobacterial Blooms in Aquatic Ecosystems Facing Increasing Anthropogenic and Climatic Pressures. Toxins, 10(2), 1–16. https://doi.org/10.3390/toxins10020076

Paerl, H. W., & Huisman, J. (2008). Climate: Blooms Like it Hot. Science, 320(5872), 57–58. https://doi.org/10.1126/science.1155398

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. Microbial Ecology, 65(4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y

Paerl, H. W., & Scott, J. T. (2010). Throwing Fuel on the Fire: Synergistic Effects of Excessive Nitrogen Inputs and Global Warming on Harmful Algal Blooms. Environmental Science and Technology, 44(20), 7756–7758. https://doi.org/10.1021/es102665e

Page, B. P., Kumar, A., & Mishra, D. R. (2018). A Novel Cross-Satellite Based Assessment of the Spatio-Temporal Development of a Cyanobacterial Harmful Algal Bloom. International Journal of Applied Earth Observation and Geoinformation, 66, 69–81. https://doi.org/10.1016/j.jag.2017.11.003

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., … Stumpf, R. (2020). Seamless Retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in Inland and Coastal Waters: A Machine-Learning Approach. Remote Sensing of Environment, 240(August 2019), 111604. https://doi.org/10.1016/j.rse.2019.111604

Palshin, N. I., Zdorovennova, G. E., Zdorovennov, R. E., Efremova, T. V., Gavrilenko, G. G., & Terzhevik, A. Y. (2019). Effect of Under-Ice Light Intensity and Convective Mixing on chlorophyll-a Distribution in a Small Mesotrophic Lake. Water Resources, 46(3), 384–394. https://doi.org/10.1134/S0097807819030175

Pan, Y., Bélanger, S., & Huot, Y. (2022). Evaluation of Atmospheric Correction Algorithms Over Lakes for High-Resolution Multispectral Imagery: Implications of Adjacency Effect. Remote Sensing, 14(13), 2979. https://doi.org/10.3390/rs14132979

Papalexiou, S. M., & Montanari, A. (2019). Global and Regional Increase of Precipitation Extremes Under Global Warming. Water Resources Research, 55(6), 4901–4914. https://doi.org/10.1029/2018WR024067

Parlee, B., & D'Souza, A. (2019). Literature Review Local and Traditional Knowledge in the Athabasca River Watershed. University of Alberta. Last Accessed: 20/03/2024, at: https://era.library.ualberta.ca/items/4aba2930-3663-4ad8-bba3-75b99069aa60

Pasquini Santos, & Dias Maciel, C. (2014). a PSO Approach for Learning Transition Structures of Higher-Order Dynamic Bayesian Networks. 5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC), 1–6. https://doi.org/10.1109/BRC.2014.6880957

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan kaufmann.

Perin, G., Buhan, I., & Picek, S. (2021). Learning When to Stop: a Mutual Information Approach to Prevent Overfitting in Profiled Side-Channel Analysis. in International Workshop on Constructive Side-Channel Analysis and Secure Design (pp. 53-81). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-89915-8_3

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing Trends in Regional Heatwaves. Nature Communications, 11(1), 1–8. https://doi.org/10.1038/s41467-020-16970-7

Perrin, B. E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., & d'Alché-Buc, F. (2003). Gene Networks Inference Using Dynamic Bayesian Networks. Bioinformatics-Oxford, 19(2), 138-148. https://doi.org/10.1093/bioinformatics/btg1071

Persaud, A. D., Paterson, A. M., Dillon, P. J., Winter, J. G., Palmer, M., & Somers, K. M. (2015). Forecasting Cyanobacteria Dominance in Canadian Temperate Lakes. Journal of Environmental Management, 151, 343–352. https://doi.org/10.1016/j.jenvman.2015.01.009

Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. Journal of Statistical Software, 27(4), 1–32. https://doi.org/10.18637/jss.v027.i04

Pick, F. R. (2016). Blooming Algae: A Canadian Perspective on the Rise of Toxic Cyanobacteria. Canadian Journal of Fisheries and Aquatic Sciences, 73(7), 1149–1158. https://doi.org/10.1139/cjfas-2015-0470

Pilla, R. M., & Williamson, C. E. (2022). Earlier Ice Breakup Induces Changepoint Responses in Duration and Variability of Spring Mixing and Summer Stratification in Dimictic Lakes. Limnology and Oceanography, 67, S173-S183. https://doi.org/10.1002/lno.11888

Pinnel, N., Heege, T., & Zimmermann, S. (2003). Spectral Discrimination of Submerged Macrophytes in Lakes Using Hyperspectral Remote Sensing Data. SPIE Proceedings Ocean Optics XVll, 16.

Poniedziałek, B., Falfushynska, H. I., & Rzymski, P. (2017). Flow Cytometry as a Valuable Tool to Study Cyanobacteria: a Mini-Review. Limnological Review, 17(2), 89-95. http://dx.doi.org/10.1515/limre-2017-0009

Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B. & Biesbroek, R. (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability: IPCC Geneva, Switzerland.

Preston, D. L., Caine, N., McKnight, D. M., Williams, M. W., Hell, K., Miller, M. P., ... & Johnson, P. T. (2016). Climate Regulates Alpine Lake Ice Cover Phenology and Aquatic Ecosystem Structure. Geophysical Research Letters, 43(10), 5353-5360. https://doi.org/10.1002/2016GL069036

Qin, B., Deng, J., Shi, K., Wang, J., Brookes, J., Zhou, J., ... & Wu, L. (2021). Extreme Climate Anomalies Enhancing Cyanobacterial Blooms in Eutrophic Lake Taihu, China. Water Resources Research, 57(7), e2020WR029371. https://doi.org/10.1029/2020WR029371

Qiu, Y., Vo, T., Garg, D., Lee, H., & Kharangate, C. R. (2023). a Systematic Approach to Optimization of ANN Model Parameters to Predict Flow Boiling Heat Transfer Coefficient in Mini/Micro-Channel Heatsinks. International Journal of Heat and Mass Transfer, 202, 123728. https://doi.org/10.1016/j.ijheatmasstransfer.2022.123728

Quesada, D. & Valverde G. (2022). DbnR: Dynamic Bayesian Network Learning and Inference. CRAN R Package. Last Accessed; 09/01/2023 at: https://github.com/dkesada/dbnR

Quesada, D., Bielza, C., & Larrañaga, P. (2021b). Structure Learning of High-Order Dynamic Bayesian Networks via Particle Swarm Optimization with Order Invariant Encoding. in Hybrid Artificial Intelligent Systems (pp. 158–171). Springer International Publishing. https://doi.org/10.1007/978-3-030-86271-8_14

Quesada, D., Valverde, G., Larrañaga, P., & Bielza, C. (2021a). Long-Term Forecasting of Multivariate Time Series in Industrial Furnaces with Dynamic Gaussian Bayesian Networks. Engineering Applications of Artificial Intelligence, 103, 104301–. https://doi.org/10.1016/j.engappai.2021.104301

Rakkasagi, S., Poonia, V., & Goyal, M. K. (2023). Flash Drought as a New Climate Threat: Drought Indices, Insights from a Study in India and Implications for Future Research. Journal of Water and Climate Change, 14(9), 3368–3384. https://doi.org/10.2166/wcc.2023.347

Rao, S., & Riahi, K. (2006). the Role of Non-CO2 Greenhouse Gases in Climate Change Mitigation: Long-Term Scenarios for the 21st Century. the Energy Journal, (Special Issue# 3).

Rao, Y. R., Huang, A., Schertzer, W. M., & Rouse, W. R. (2012). Modelling of Physical Processes and Assessment of Climate Change Impacts in Great Bear Lake. Atmosphere-Ocean, 50(3), 317–333. https://doi.org/10.1080/07055900.2012.668492

Rasouli, K., Hernández-Henríquez, M. A., & Déry, S. J. (2013). Streamflow Input to Lake Athabasca, Canada. Hydrology and Earth System Sciences, 17(5), 1681-1691. https://doi.org/10.5194/hess-17-1681-2013

Rasti, B., Scheunders, P., Ghamisi, P., Licciardi, G., & Chanussot, J. (2018). Noise Reduction in Hyperspectral Imagery: Overview and Application. Remote Sensing, 10(3), 1–28. https://doi.org/10.3390/rs10030482

Raven, J. A. (1988). The Iron and Molybdenum Use Efficiencies of Plant Growth with Different Energy, Carbon and Nitrogen Sources. New Phytologist, 109(3), 279–287. https://doi.org/10.1111/j.1469-8137.1988.tb04196.x

Raven, J. A. (1990). Predictions of Mn and Fe Use Efficiencies of Phototrophic Growth as a Function of Light Availability for Growth and of C Assimilation Pathway. New Phytologist, 116(1), 1–18. https://doi.org/10.1111/j.1469-8137.1990.tb00505.x

Raven, J. A., Gobler, C. J., & Hansen, P. J. (2020). Dynamic CO2 and Ph Levels in Coastal, Estuarine, and Inland Waters: Theoretical and Observed Effects on Harmful Algal Blooms. Harmful Algae, 91(April 2019), 101594. https://doi.org/10.1016/j.hal.2019.03.012

Reichwaldt, E. S., & Ghadouani, A. (2012). Effects of Rainfall Patterns on Toxic Cyanobacterial Blooms in a Changing Climate: Between Simplistic Scenarios and Complex Dynamics. Water Research, 46(5), 1372–1393. https://doi.org/10.1016/j.watres.2011.11.052

Reifsnyder. W. E. (1967). Radiation Geometry in the Measurement and Interpretation of Radiation Balance. Agricultural Meteorology, 4(4), 255–265. https://doi.org/10.1016/0002-1571(67)90026-X

Reinl, K. L., Brookes, J. D., Carey, C. C., Harris, T. D., Ibelings, B. W., Morales-Williams, A. M., De Senerpont Domis, L. N., Atkins, K. S., Isles, P. D. F., Mesman, J. P., North, R. L., Rudstam, L. G., Stelzer, J. A. A., Venkiteswaran, J. J., Yokota, K., & Zhan, Q. (2021). Cyanobacterial Blooms in Oligotrophic Lakes: Shifting the High-Nutrient Paradigm. Freshwater Biology, 66(9), 1846–1859. https://doi.org/10.1111/fwb.13791

Reynolds, C. S. (1994). the Long, the Short and the Stalled: on the Attributes of Phytoplankton Selected by Physical Mixing in Lakes and Rivers. Phytoplankton in Turbid Environments: Rivers and Shallow Lakes, 9–21. https://doi.org/10.1007/978-94-017-2670-2_2

Reynolds, C. S., & Walsby, A. E. (1975). Water-blooms. Biological reviews, 50(4), 437-481. https://doi.org/10.1111/j.1469-185X.1975.tb01060.x

Reynolds, G. T., & Lutz, R. A. (2001). Sources of Light in the Deep Ocean. Reviews of Geophysics, 39(1), 123–136. https://doi.org/10.1029/1999RG000071

Rhodes. T. A. (2017). An Evaluation of Critical Realignment Theory: Comparing Bayesian and Frequentist Approaches. ProQuest Dissertations Publishing.

Rigosi, A., Hanson, P., Hamilton, D. P., Hipsey, M., Rusak, J. A., Bois, J., ... & Brookes, J. D. (2015). Determining the Probability of Cyanobacterial Blooms: The Application of Bayesian Networks in Multiple Lake Systems. Ecological Applications, 25(1), 186-199. https://doi.org/10.1890/13-1677.1

Ritchie, J. C., Cooper, C. M., & Schiebe, F. R. (1990). the Relationship of MSS and TM Digital Data with Suspended Sediments, Chlorophyll, and Temperature in Moon Lake, Mississippi. Remote Sensing of Environment, 33(2), 137–148. https://doi.org/10.1016/0034-4257(90)90039-O

Roberts, J. J., Fausch, K. D., Schmidt, T. S., & Walters, D. M. (2017). Thermal Regimes of Rocky Mountain Lakes Warm with Climate Change. PLoS ONE, 12(7), 1–17. https://doi.org/10.1371/journal.pone.0179498

Robertson, D. M., & Diebel, M. W. (2020). Importance of Accurately Quantifying Internal Loading in Developing Phosphorus Reduction Strategies for a Chain of Shallow Lakes. Lake and Reservoir Management, 36(4), 391-411. https://doi.org/10.1080/10402381.2020.1783727

Robinson, J. W., Hartemink, A. J., & Ghahramani, Z. (2010). Learning Non-Stationary Dynamic Bayesian Networks. Journal of Machine Learning Research, 11(12).

Rocha, S. M. G., da Silva, J. V. B., Lemos, W. E. D., de Souza Filho, F. de A., & Lima Neto, I. E. (2022). Two-Dimensional Modelling of the Mixing Patterns in a Tropical Semiarid Reservoir. Sustainability (Basel, Switzerland), 14(23), 16051–. https://doi.org/10.3390/su142316051

Rodrigues, G., Potes, M., Penha, A. M., Costa, M. J., & Morais, M. M. (2022). the Use of Sentinel-3/OLCI for Monitoring the Water Quality and Optical Water Types in the Largest Portuguese Reservoir. Remote Sensing (Basel, Switzerland), 14(9), 2172–. https://doi.org/10.3390/rs14092172

Rodríguez-López, L., Duran-Llacer, I., Bravo Alvarez, L., Lami, A., & Urrutia, R. (2023). Recovery of Water Quality and Detection of Algal Blooms in Lake Villarrica through Landsat Satellite Images and Monitoring Data. Remote Sensing (Basel, Switzerland), 15(7), 1929–. https://doi.org/10.3390/rs15071929

Roesler, C. (2014). Lecture 1: Bio-optical modeling of IOPs (PFT approaches). International Ocean Colour Coordinating Group (IOCCG) Summer Lecture Series 2014. Last accessed 08/26/2020 at: https://www.ioccg.org/training/SLS-2014/Roesler_Semi-analytical_inversion_2014_IOCCG.pdf

Roesler, C. S., & Perry, M. J. (1995). in Situ Phytoplankton Absorption, Fluorescence Emission, and Particulate Backscattering Spectra Determined from Reflectance. Journal of Geophysical Research: Oceans, 100(C7), 13279-13294.

Roesler, C. S., Perry, M. J., & Carder, K. L. (1989). Modeling in Situ Phytoplankton Absorption from Total Absorption Spectra in Productive Inland Marine Waters. Limnology and Oceanography, 34(8), 1510–1523. https://doi.org/10.4319/lo.1989.34.8.1510

Rohwer, R. R., Hale, R. J., Vander Zanden, M. J., Miller, T. R., & McMahon, K. D. (2023). Species Invasions Shift Microbial Phenology in a Two-Decade Freshwater Time Series. Proceedings of the National Academy of Sciences - PNAS, 120(11), e2211796120–e2211796120. https://doi.org/10.1073/pnas.2211796120

Roman, M. R., Brandt, S. B., Houde, E. D., & Pierson, J. J. (2019). Interactive Effects of Hypoxia and Temperature on Coastal Pelagic Zooplankton and Fish. Frontiers in Marine Science, 6(MAR), 1–18. https://doi.org/10.3389/fmars.2019.00139

Rose, K. C., Winslow, L. A., Read, J. S., & Hansen, G. J. A. (2016). Climate-Induced Warming of Lakes Can be Either Amplified or Suppressed by Trends in Water Clarity. Limnology and Oceanography Letters, 1(1), 44–53. https://doi.org/10.1002/lol2.10027

Rouse, Blanken, P. D., Duguay, C. R., Oswald, C. J., & Schertzer, W. M. (2008). Climate-Lake Interactions. in Cold Region Atmospheric and Hydrologic Studies. The Mackenzie GEWEX Experience (Vol. 2, pp. 139–160). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75136-6_8

Rühland, K. M., Evans, M., & Smol, J. P. (2023). Arctic Warming Drives Striking Twenty-First Century Ecosystem Shifts in Great Slave Lake (Subarctic Canada), North America's Deepest Lake. Proceedings of the Royal Society. B, Biological Sciences, 290(2007), 20231252–20231252. https://doi.org/10.1098/rspb.2023.1252

Ruiz-Tagle, A., Lewis, A. D., Schell, C. A., Lever, E., & Groth, K. M. (2022). BaNTERA: a Bayesian Network for Third-Party Excavation Risk Assessment. Reliability Engineering & System Safety, 223, 108507. https://doi.org/10.1016/j.ress.2022.108507

Ruiz-Verdú, A., Simis, S. G. H., de Hoyos, C., Gons, H. J., & Peña-Martínez, R. (2008). an Evaluation of Algorithms for the Remote Sensing of Cyanobacterial Biomass. Remote Sensing of Environment, 112(11), 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., … Adams, C. (2020). Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-Optical Simulations, Machine Learning, and Cloud Computing. Earth-Science Reviews, 205(April), 103187. https://doi.org/10.1016/j.earscirev.2020.103187

Salem, S. I., Higa, H., Kim, H., Kazuhiro, K., Kobayashi, H., Oki, K., & Oki, T. (2017). Multi-Algorithm Indices and Look-Up Table for chlorophyll-a Retrieval in Highly Turbid Water Bodies Using Multispectral Data. Remote Sensing (Basel, Switzerland), 9(6), 556–. https://doi.org/10.3390/rs9060556

Salmaso, N. (2010). Long-Term Phytoplankton Community Changes in a Deep Subalpine Lake: Responses to Nutrient Availability and Climatic Fluctuations. Freshwater Biology, 55(4), 825–846. https://doi.org/10.1111/j.1365-2427.2009.02325.x

Salmaso, N., Boscaini, A., Capelli, C., & Cerasino, L. (2018). Ongoing Ecological Shifts in a Large Lake are Driven by Climate Change and Eutrophication: Evidences from a Three-Decade Study in Lake Garda. Hydrobiologia, 824(1), 177–195. https://doi.org/10.1007/s10750-017-3402-1

Samaniego, F. J. (2010). A Comparison of the Bayesian and Frequentist Approaches to Estimation (Vol. 24). New York: Springer.

Samaniego, L., Thober, S., Kumar, R., Wanders, N., Rakovec, O., Pan, M., … Marx, A. (2018). Anthropogenic Warming Exacerbates European Soil Moisture Droughts. Nature Climate Change, 8(5), 421–426. https://doi.org/10.1038/s41558-018-0138-5

Santos, F. P., Smagula, S. F., Karim, H. T., Santini, T. S., Aizenstein, H., Ibrahim, T. S., & Maciel, C. D. (2017, January). Dynamic Bayesian Network Modeling of Hippocampal Subfields Connectivity with 7T fMRI: a Case Study. in BIOSIGNALS (pp. 178-184).

Sapriza-Azuri, G., Gamazo, P., Razavi, S., & Wheater, H. S. (2018). on the Appropriate Definition of Soil Profile Configuration and Initial Conditions for Land Surface–Hydrology Models in Cold Regions. Hydrology and Earth System Sciences, 22(6), 3295–3309. https://doi.org/10.5194/hess-22-3295-2018

Sass, G. Z., Creed, I. F., Bayley, S. E., & Devito, K. J. (2007). Understanding Variation in Trophic Status of Lakes on the Boreal Plain: a 20 Year Retrospective Using Landsat TM Imagery. Remote Sensing of Environment, 109(2), 127–141. https://doi.org/10.1016/j.rse.2006.12.010

Sayers, M. J., Grimm, A. G., Shuchman, R. A., Bosse, K. R., Fahnenstiel, G. L., Ruberg, S. A., & Leshkevich, G. A. (2019). Satellite Monitoring of Harmful Algal Blooms in the Western Basin of Lake Erie: a 20-Year Time-Series. Journal of Great Lakes Research, 45(3), 508–521. https://doi.org/10.1016/j.jglr.2019.01.005

Schalles, J. F. (2006). Optical Remote Sensing Techniques to Estimate Phytoplankton Chlorophyll-aConcentrations in Coastal. In:Remote Sensing of Aquatic Coastal Ecosystem Processes. L. Richardson & E.LeDrew (eds.), Springer, Dordrecht.pp. 27-79. https://doi.org/10.1007/1-4020-3968-9_3

Schertzer, W. M., Rouse, W. R., & Blanken, P. D. (2000). Cross-Lake Variation of Physical Limnological and Climatological Processes of Great Slave Lake. Physical Geography, 21(5), 385–406. https://doi.org/10.1080/02723646.2000.10642716

Schertzer, W. M., Rouse, W. R., Blanken, P. D., Walker, A. E., Lam, D. C., & León, L. (2008). Interannual Variability of the Thermal Components and Bulk Heat Exchange of Great Slave Lake. Cold Region Atmospheric and Hydrologic Studies. the Mackenzie GEWEX Experience: Volume 2: Hydrologic Processes, 197-219. https://doi.org/10.1007/978-3-540-75136-6_11

Schindler, D. W., Carpenter, S. R., Chapra, S. C., Hecky, R. E., & Orihel, D. M. (2016). Reducing Phosphorus to Curb Lake Eutrophication is a Success. Environmental Science and Technology, 50(17), 8923–8929. https://doi.org/10.1021/acs.est.6b02204

Schindler, D. W., Hecky, R. E., Findlay, D. L., Stainton, M. P., Parker, B. R., Paterson, M. J., … Kasian, S. E. M. (2008). Eutrophication of Lakes Cannot be Controlled by Reducing Nitrogen Input: Results of a 37-Year Whole-Ecosystem Experiment. Proceedings of the National Academy of Sciences of the United States of America, 105(32), 11254–11258. https://doi.org/10.1073/pnas.0805108105

Schindler. D. W. (1977). Evolution of Phosphorus Limitation in Lakes. Science (American Association for the Advancement of Science), 195(4275), 260–262. https://doi.org/10.1126/science.195.4275.260

Schindler. D. W. (2001). the Cumulative Effects of Climate Warming and Other Human Stresses on Canadian Freshwaters in the New Millennium. Canadian Journal of Fisheries and Aquatic Sciences, 58(1), 18–29. https://doi.org/10.1139/f00-179

Schindler. D.W. (1974). Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management. Science (American Association for the Advancement of Science), 184(4139), 897–899. https://doi.org/10.1126/science.184.4139.897

Schlabing, D., Frassl, M. A., Eder, M. M., Rinke, K., & Bárdossy, A. (2014). Use of a Weather Generator for Simulating Climate Change Effects on Ecosystems: A Case Study on Lake Constance. Environmental modelling & software, 61, 326-338. https://doi.org/10.1016/j.envsoft.2014.06.028

Schlegel, R. W., Oliver, E. C. J., Hobday, A. J., & Smit, A. J. (2019). Detecting Marine Heatwaves with Sub-Optimal Data. Frontiers in Marine Science, 6. https://doi.org/10.3389/fmars.2019.00737

Schott, J. R., Gerace, A., Woodcock, C. E., Wang, S., Zhu, Z., Wynne, R. H., & Blinn, C. E. (2016). The Impact of Improved Signal-To-Noise Ratios on Algorithm Performance: Case Studies for Landsat Class Instruments. Remote Sensing of Environment, 185, 37–45. https://doi.org/10.1016/j.rse.2016.04.015

Screen, J. A., & Simmonds, I. (2010). The Central Role of Diminishing Sea Ice in Recent Arctic Temperature Amplification. Nature (London), 464(7293), 1334–1337. https://doi.org/10.1038/nature09051

Scutari, M., Scutari, M. M., & MMPC, H. P. (2019). Package 'bnlearn'. Bayesian Network Structure Learning, Parameter Learning and Inference, R package version, 4.8.3.

Sebastiani, P., Abad, M., & Ramoni, M. F. (2005). Bayesian Networks for Genomic Analysis. Genomic Signal Processing and Statistics, 2, 281-320.

Sebastiani, P., Abad, M.M., Ramoni, M.F. (2009). Bayesian Networks. In: Maimon, O., Rokach, L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-09823-4_10

Seekell, D. A., Carpenter, S. R., & Pace, M. L. (2011). Conditional Heteroscedasticity as a Leading Indicator of Ecological Regime Shifts. the American Naturalist, 178(4), 442-451. https://doi.org/10.1086/661898

Senar, O. E., Creed, I. F., Strandberg, U., & Arts, M. T. (2019). Browning Reduces the Availability—But Not the Transfer—Of Essential Fatty Acids in Temperate Lakes. Freshwater Biology, 64(12), 2107–2119. https://doi.org/10.1111/fwb.13399

Sforza, E., Simionato, D., Giacometti, G. M., Bertucco, A., & Morosinotto, T. (2012). Adjusted Light and Dark Cycles Can Optimize Photosynthetic Efficiency in Algae Growing in Photobioreactors. PloS One, 7(6), e38975–e38975. https://doi.org/10.1371/journal.pone.0038975

Shan, K., Shang, M., Zhou, B., Li, L., Wang, X., Yang, H., & Song, L. (2019). Application of Bayesian Network Including Microcystis Morphospecies for Microcystin Risk Assessment in Three Cyanobacterial Bloom-Plagued Lakes, China. Harmful Algae, 83, 14-24. https://doi.org/10.1016/j.hal.2019.01.005

Shanmuga Priyaa, S., Aruna Kumar, A., & Jena, B. K. (2022). Bathymetry Retrieval Using Remote Sensing Techniques for Inter-tidal Regions of Tapi Estuary. in River and Coastal Engineering: Hydraulics, Water Resources and Coastal Engineering (pp. 213-226). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-05057-2_19

Sharma, S., Blagrave, K., Magnuson, J. J., O'Reilly, C. M., Oliver, S., Batt, R. D., … Woolway, R. I. (2019). Widespread Loss of Lake Ice Around the Northern Hemisphere in a Warming World. Nature Climate Change, 9(3), 227–231. https://doi.org/10.1038/s41558-018-0393-5

Sharma, S., Magnuson, J. J., Mendoza, G., & Carpenter, S. R. (2013). Influences of Local Weather, Large-Scale Climatic Drivers, and the Ca. 11 Year Solar Cycle on Lake Ice Breakup Dates; 1905-2004. Climatic Change, 118(3–4), 857–870. https://doi.org/10.1007/s10584-012-0670-7

Sharma, S., Richardson, D. C., Woolway, R. I., Imrit, M. A., Bouffard, D., Blagrave, K., Daly, J., Filazzola, A., Granin, N., Korhonen, J., Magnuson, J., Marszelewski, W., Matsuzaki, S. S., Perry, W., Robertson, D. M., Rudstam, L. G., Weyhenmeyer, G. A., & Yao, H. (2021). Loss of Ice Cover, Shifting Phenology, and More Extreme Events in Northern Hemisphere Lakes. Journal of Geophysical Research. Biogeosciences, 126(10). https://doi.org/10.1029/2021JG006348

Shi, X., Gu, L., Jiang, T., Zheng, X., Dong, W., & Tao, Z. (2022). Retrieval of chlorophyll-a Concentrations Using Sentinel-2 MSI Imagery in Lake Chagan Based on Assessments with Machine Learning Models. Remote Sensing, 14(19), 4924. https://doi.org/10.3390/rs14194924

Shi, Z., Krom, M. D., Jickells, T. D., Bonneville, S., Carslaw, K. S., Mihalopoulos, N., … Benning, L. G. (2012). Impacts on Iron Solubility in the Mineral Dust by Processes in the Source Region and the Atmosphere: a Review. Aeolian Research, 5(May 2018), 21–42. https://doi.org/10.1016/j.aeolia.2012.03.001

Siegel, D. A., Wang, M., Maritorena, S., & Robinson, W. (2000). Atmospheric Correction of Satellite Ocean Color Imagery: The Black Pixel Assumption. Applied Optics, 39(21), 3582. https://doi.org/10.1364/ao.39.003582

Simis, S. G. H., Calmettes B., Crétaux, J. F., Liu, X., Mangili, H., Yésou, L., Duguay, C. R., Wu, Y. (2023) Lakes_CCI+ - Phase 2 – D2.1. Product Validation and Algorithm Selection Report (PVASR). European Space Agency. Last Accessed: 03/22/2024, at: https://climate.esa.int/media/documents/Lakes_CCI_D2-1_PVASR.pdf

Simis, S. G. H., Peters, S. W. M., & Gons, H. J. (2005). Remote Sensing of the Cyanobacterial Pigment Phycocyanin in Turbid Inland Water. Limnology and Oceanography, 50(1), 237–245. https://doi.org/10.4319/lo.2005.50.1.0237

Singh, K., Ghosh, M., Sharma, S. R., & Kumar, P. (2014). Blue-Red-NIR Model for Chlorophyll-A Retrieval in Hypersaline-Alkaline Water Using Landsat ETM+ Sensor. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(8), 3553-3559. https://doi.org/10.1109/JSTARS.2014.2340856

Singh, S. P., & Singh, P. (2015). Effect of Temperature and Light on the Growth of Algae Species: a Review. Renewable & Sustainable Energy Reviews, 50, 431–444. https://doi.org/10.1016/j.rser.2015.05.024

Sivarajah, B., Simmatis, B., Favot, E. J., Palmer, M. J., & Smol, J. P. (2021). Eutrophication and Climatic Changes Lead to Unprecedented Cyanobacterial Blooms in a Canadian Sub-Arctic Landscape. Harmful Algae, 105, 102036–102036. https://doi.org/10.1016/j.hal.2021.102036

Skwierawski, A. (2022). Carbon Sequestration Potential in the Restoration of Highly Eutrophic Shallow Lakes. International Journal of Environmental Research and Public Health, 19(10), 6308–. https://doi.org/10.3390/ijerph19106308

Smayda, T. J. (2008). Complexity in the Eutrophication-Harmful Algal Bloom Relationship, with Comment on the Importance of Grazing. Harmful Algae, 8(1), 140–151. https://doi.org/10.1016/j.hal.2008.08.018

Šmejkalová, T., Edwards, M. E., & Dash, J. (2016). Arctic Lakes Show Strong Decadal Trend in Earlier Spring Ice-Out. Scientific Reports, 6(December), 1–8. https://doi.org/10.1038/srep38449

Smith, Bass, B., Sawyer, D., Depew, D., & Watson, S. B. (2019). Estimating the Economic Costs of Algal Blooms in the Canadian Lake Erie Basin. Harmful Algae, 87, 101624–101624. https://doi.org/10.1016/j.hal.2019.101624

Smith, D. R., King, K. W., & Williams, M. R. (2015). What is Causing the Harmful Algal Blooms in Lake Erie? Journal of Soil and Water Conservation, 70(2), 27A-29A. https://doi.org/10.2489/jswc.70.2.27A

Smith, J. A., Baeck, M. L., Villarini, G., Wright, D. B., & Krajewski, W. (2013). Extreme Flood Response: the June 2008 Flooding in Iowa. Journal of Hydrometeorology, 14(6), 1810–1825. https://doi.org/10.1175/JHM-D-12-0191.1

Smith, R. C., & Baker, K. S. (1978). the Bio-Optical State of Ocean Waters and Remote Sensing. Limnology and Oceanography, 23(2), 247–259. https://doi.org/10.4319/lo.1978.23.2.0247

Smol. J, P. (2019). Under the Radar: Long-Term Perspectives on Ecological Changes in Lakes. Proceedings of the Royal Society. B, Biological Sciences, 286(1906), 20190834–20190834. https://doi.org/10.1098/rspb.2019.0834

Soja-Woźniak, M., Laiolo, L., Baird, M. E., Matear, R., Clementson, L., Schroeder, T., … Suthers, I. M. (2020). Effect of Phytoplankton Community Size Structure on Remote-Sensing Reflectance and chlorophyll-a Products. Journal of Marine Systems, 211(July). https://doi.org/10.1016/j.jmarsys.2020.103400

Sommer, U., Adrian, R., De Senerpont Domis, L., Elser, J. J., Gaedke, U., Ibelings, B., … Winder, M. (2012). Beyond the Plankton Ecology Group (PEG) Model: Mechanisms Driving Plankton Succession. Annual Review of Ecology, Evolution, and Systematics, 43(1), 429–448. https://doi.org/10.1146/annurev-ecolsys-110411-160251

Song, K. (2011). Water Quality Monitoring Using Landsat Themate Mapper Data with Empirical Algorithms in Chagan Lake, China. Journal of Applied Remote Sensing, 5(1), 053506. https://doi.org/10.1117/1.3559497

Song, L., Qin, J. G., Su, S., Xu, J., Clarke, S., & Shan, Y. (2012). Micronutrient Requirements for Growth and Hydrocarbon Production in the Oil Producing Green Alga Botryococcus Braunii (Chlorophyta). PLoS ONE, 7(7), 1–9. https://doi.org/10.1371/journal.pone.0041459

Song, Z., Liang, S., & Zhou, H. (2021). Top-of-Atmosphere Clear-Sky Albedo Estimation Over Ocean: Preliminary Framework for MODIS. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-9. https://doi.org/10.1109/TGRS.2021.3116620

Spyrakos, E., O'Donnell, R., Hunter, P. D., Miller, C., Scott, M., Simis, S. G. H., Neil, C., Barbosa, C. C. F., Binding, C. E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A. A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M. W., … Tyler, A. N. (2018). Optical Types of Inland and Coastal Waters. Limnology and Oceanography, 63(2), 846–870. https://doi.org/10.1002/lno.10674

StataCorp, L. P. (2023). Bayesian Analysis Reference Manual, Release 18. College Station, TX: StataCorp LP. Last Accessed; 09/01/2023, at: https://www.stata.com/manuals/bayes.pdf

Statistics Canada (2016). Census profile, 2016 census: Fort Chipewyan, Unincorporated place [Designated place], Alberta and Alberta [Province]. Statistics Canada. Last Accessed: 4/20/2022 at https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=DPL&Code1=480243&Geo2=PR&Code2=48&SearchText=Fort%20Chipewyan&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=480243&TABID=1&type=0

Stefanidis, K., Varlas, G., Papaioannou, G., Papadopoulos, A., & Dimitriou, E. (2022). Trends of Lake Temperature, Mixing Depth and Ice Cover Thickness of European Lakes During the Last Four Decades. Science of the Total Environment, 830, 154709. https://doi.org/10.1016/j.scitotenv.2022.154709

Stockenreiter, M., Isanta Navarro, J., Buchberger, F., & Stibor, H. (2021). Community Shifts from Eukaryote to Cyanobacteria Dominated Phytoplankton: The Role of Mixing Depth and Light Quality. Freshwater Biology, 66(11), 2145–2157. https://doi.org/10.1111/fwb.13822

Stuart, M. B., McGonigle, A. J. S., & Willmott, J. R. (2019). Hyperspectral Imaging in Environmental Monitoring: a Review of Recent Developments and Technological Advances in Compact Field Deployable Systems. Sensors (Switzerland), 19(14). https://doi.org/10.3390/s19143071

Su, C., Andrew, A., Karagas, M. R., & Borsuk, M. E. (2013). Using Bayesian Networks to Discover Relations Between Genes, Environment, and Disease. BioData mining, 6(1), 1-21. https://doi.org/10.1186/1756-0381-6-6

Sudheer, K. P., Chaubey, I., & Garg, V. (2006). Lake Water Quality Assessment from Landsat Thematic Mapper Data Using Neural Network: An Approach to Optimal Band Combination Selection. Journal of the American Water Resources Association, 42(6), 1683–1695. https://doi.org/10.1111/j.1752-1688.2006.tb06029.x

Suter, P., Kuipers, J., & Beerenwinkel, N. (2022). Discovering Gene Regulatory Networks of Multiple Phenotypic Groups Using Dynamic Bayesian Networks. Briefings in Bioinformatics, 23(4), bbac219. https://doi.org/10.1093/bib/bbac219

Syariz, M. A., Lin, C. H., & Blanco, A. C. (2019). chlorophyll-a Concentration Retrieval Using Convolutional Neural Networks in Laguna Lake, Philippines. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(4/W19), 401–405. https://doi.org/10.5194/isprs-archives-XLII-4-W19-401-2019

Talbot. (1990). A Review of the Palaeohydrological Interpretation of Carbon and Oxygen Isotopic Ratios in Primary Lacustrine Carbonates. Chemical Geology: Isotope Geoscience Section, 80(4), 261–279. https://doi.org/10.1016/0168-9622(90)90009-2

Talling, J. F. (1976). the Depletion of Carbon Dioxide from Lake Water by Phytoplankton. the Journal of Ecology, 79-121. https://doi.org/10.2307/2258685

Tammeorg, O., Nürnberg, G., Niemistö, J. et al. (2020). Internal Phosphorus Loading Due to Sediment Anoxia in Shallow Areas: Implications for Lake Aeration Treatments. Aquat Sci 82, 54. https://doi.org/10.1007/s00027-020-00724-0

Tanentzap, A. J., Yan, N. D., Keller, B., Girard, R., Heneberry, J., Gunn, J. M., … Taylor, P. A. (2008). Cooling Lakes While the World Warms: Effects of Forest Regrowth and Increased Dissolved Organic Matter on the Thermal Regime of a Temperate, Urban Lake. Limnology and Oceanography, 53(1), 404–410. https://doi.org/10.4319/lo.2008.53.1.0404

Tanré, D., Deroo, C., Duhaut, P., Herman, M., Morcrette, J. J., Perbos, J., & Deschamps, P. Y. (1990). Description of a Computer Code to Simulate the Satellite Signal in the Solar Spectrum: the 5S code. International Journal of Remote Sensing, 11(4), 659–668. https://doi.org/10.1080/01431169008955048

Tao, Y., Yu, J., Xue, B., Yao, S., & Wang, S. (2017). Precipitation and Temperature Drive Seasonal Variation in Bioaccumulation of Polycyclic Aromatic Hydrocarbons in the Planktonic Food Webs of a Subtropical Shallow Eutrophic Lake in China. Science of the Total Environment, 583, 447-457. https://doi.org/10.1016/j.scitotenv.2017.01.100

Thackeray, C. W., Hall, A., Norris, J., & Chen, D. (2022). Constraining the Increased Frequency of Global Precipitation Extremes Under Warming. Nature Climate Change, 12(5), 441–448. https://doi.org/10.1038/s41558-022-01329-1

Thomas, M. K., & Litchman, E. (2016). Effects of Temperature and Nitrogen Availability on the Growth of Invasive and Native Cyanobacteria. Hydrobiologia, 763(1), 357–369. https://doi.org/10.1007/s10750-015-2390-2

Tian, D., Xie, G., Tian, J., Tseng, K. H., Shum, C. K., Lee, J., & Liang, S. (2017). Spatiotemporal Variability and Environmental Factors of Harmful Algal Blooms (Habs) Over Western Lake Erie. PloS one, 12(6), e0179622. https://doi.org/10.1371/journal.pone.0179622

Tieleman, T., & Hinton, G. (2017). Divide the Gradient by a Running Average of Its Recent Magnitude. Coursera: Neural Networks for Machine Learning. Technical Report.

Timoney, K. (2008). a Study of Water and Sediment Quality as Related to Public Health Issues, Fort Chipewyan Alberta. Sherwood Park: Nunee Health Board Society, Fort Chipewyan Alberta.

Tiwari, S. P., Adhikary, S., & Banerjee, S. (2022). Estimation of chlorophyll-a from Oceanographic Properties-An Indirect Approach. in IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 6872-6875). IEEE. https://doi.org/10.1109/IGARSS46834.2022.9883294

Tong, Y., Feng, L., Zhao, D., Xu, W., & Zheng, C. (2022). Remote Sensing of chlorophyll-a Concentrations in Coastal Oceans of the Greater Bay Area in China: Algorithm Development and Long-Term Changes. International Journal of Applied Earth Observation and Geoinformation, 112, 102922–. https://doi.org/10.1016/j.jag.2022.102922

Trainer, V. M., Moore, S. K., Hallegraeff, G., Kudela, R. M., Clement, A., Mardones, J. I., & Cochlan, W. P. (2020). Pelagic Harmful Algal Blooms and Climate Change: Lessons from Nature's Experiments with Extremes. Harmful Algae, 91, 101591–101591. https://doi.org/10.1016/j.hal.2019.03.009

Trochine, C., Guerrieri, M., Liboriussen, L., Meerhoff, M., Lauridsen, T. L., Søndergaard, M., & Jeppesen, E. (2011). Filamentous Green Algae Inhibit Phytoplankton with Enhanced Effects When Lakes Get Warmer. Freshwater Biology, 56(3), 541-553. https://doi.org/10.1111/j.1365-2427.2010.02521.x

Trofimova, Y., Fesl, J., & Moucha, A. M. (2021, September). Performance Analysis of Neural Network Approach for Evaluation of Trust in Ad-Hoc Networks. in 2021 11th International Conference on Advanced Computer Information Technologies (ACIT) (pp. 691-695). IEEE. https://doi.org/10.1109/ACIT52158.2021.9548546

Tsvetova. E. A. (2016). Natural and Forced Under-Ice Convection. Proceedings of SPIE - the International Society for Optical Engineering, 10035, 1003568–1003568–5. https://doi.org/10.1117/12.2249022

Valero-Leal, E., Larranaga, P., & Bielza, C. (2022). Interpreting Time-Varying Dynamic Bayesian Networks for Earth Climate Modelling. in International Conference on Probabilistic Graphical Models (pp. 373-384). PMLR.

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing Off with Scylla and Charybdis: a Comparison of Scalar, Partial, and the Novel Possibility of Approximate Measurement Invariance. Frontiers in Psychology, 4, 770–770. https://doi.org/10.3389/fpsyg.2013.00770

Varotsos, C. A., Efstathiou, M. N., & Cracknell, A. P. (2013). on the Scaling Effect in Global Surface Air Temperature Anomalies. Atmospheric Chemistry and Physics, 13(10), 5243–5253. https://doi.org/10.5194/acp-13-5243-2013

Vermote, E. F., Tanré, D., Deuzé, J. L., Herman, M., & Morcrette, J. J. (1997). Second Simulation of the Satellite Signal in the Solar Spectrum, 6s: an Overview. IEEE Transactions on Geoscience and Remote Sensing, 35(3), 675–686. https://doi.org/10.1109/36.581987

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A Global Inventory of Lakes Based on High-Resolution Satellite Imagery. Geophysical Research Letters, 41(18), 6396–6402. https://doi.org/10.1002/2014GL060641

Verschoor, A. M., Van Dijk, M. A., Huisman, J., & Van Donk, E. (2013). Elevated CO2 Concentrations Affect the Elemental Stoichiometry and Species Composition of an Experimental Phytoplankton Community. Freshwater Biology, 58(3), 597–611. https://doi.org/10.1111/j.1365-2427.2012.02833.x

Vignes, M., Vandel, J., Allouche, D., Ramadan-Alban, N., Cierco-Ayrolles, C., Schiex, T., ... & De Givry, S. (2011). Gene Regulatory Network Reconstruction Using Bayesian Networks, the Dantzig Selector, the Lasso and Their Meta-Analysis. PloS one, 6(12), e29165. https://doi.org/10.1371/journal.pone.0029165

Vilhena, L. C., Hillmer, I., & Imberger, J. R. (2010). the Role of Climate Change in the Occurrence of Algal Blooms: Lake Burragorang, Australia. Limnology and Oceanography, 55(3), 1188-1200. https://doi.org/10.4319/lo.2010.55.3.1188

Vincent, R. K. (1972). An ERTS Multispectral Scanner Experiment for Mapping Iron Compounds. NASA Technical Report Server. Last Accessed (08/05/2020) at: https://ntrs.nasa.gov/search.jsp?R=19730001633

Vincent, R. K., Qin, X., McKay, R. M. L., Miner, J., Czajkowski, K., Savino, J., & Bridgeman, T. (2004). Phycocyanin Detection from LANDSAT TM Data for Mapping Cyanobacterial Blooms in Lake Erie. Remote Sensing of Environment, 89(3), 381–392. https://doi.org/10.1016/j.rse.2003.10.014

Vincent, W. F., Laurion, I., Pienitz, R., & Walter Anthony, K. M. (2012). Climate Impacts on Arctic Lake Ecosystems. Climatic Change and Global Warming of Inland Waters: Impacts and Mitigation for Ecosystems and Societies, (2011), 27–42. https://doi.org/10.1002/9781118470596.ch2

Vitale, V., Musella, F., Vicard, P., & Guizzi, V. (2020). Modelling an Energy Market with Bayesian Networks for Non-Normal Data. Computational Management Science, 17(1), 47–64. https://doi.org/10.1007/s10287-018-0320-2

Walsh, J. R., Hansen, G. J. A., Read, J. S., & Vander Zanden, M. J. (2020). Comparing Models Using Air and Water Temperature to Forecast an Aquatic Invasive Species Response to Climate Change. Ecosphere (Washington, D.C), 11(7). https://doi.org/10.1002/ecs2.3137

Wang, D., Ma, R., Xue, K., & Loiselle, S. A. (2019). The Assessment of Landsat-8 OLI Atmospheric Correction Algorithms for Inland Waters. Remote Sensing, 11(2), 169. https://doi.org/10.3390/rs11020169

Wang, D., Xiang, X., Ma, R., Guo, Y., Zhu, W., & Wu, Z. (2023). A Novel Atmospheric Correction for Turbid Water Remote Sensing. Remote Sensing (Basel, Switzerland), 15(8), 2091–. https://doi.org/10.3390/rs15082091

Wang, K. J., Chen, J. L., Chen, K. H., & Wang, K. M. (2020). Survivability Prognosis for Lung Cancer Patients at Different Severity Stages by a Risk Factor-Based Bayesian

Network Modeling. Journal of Medical Systems, 44(3), 65. https://doi.org/10.1007/s10916-020-1537-5

Wang, Li, C., Xu, Y.-P., Li, S.-Y., Du, J.-S., Han, Y.-P., & Hu, H.-Y. (2021). Identifying Major Contributors to Algal Blooms in Lake Dianchi by Analyzing River-Lake Water Quality Correlations in the Watershed. Journal of Cleaner Production, 315, 128144–. https://doi.org/10.1016/j.jclepro.2021.128144

Wang, T. S., Tan, C. H., Chen, L., & Tsai, Y. C. (2008). Applying Artificial Neural Networks and Remote Sensing to Estimate chlorophyll-a Concentration in Water Body. Proceedings - 2008 2nd International Symposium on Intelligent Information Technology Application, IITA 2008, 1, 540–544. https://doi.org/10.1109/IITA.2008.279

Wang, X., Shi, K., Zhang, Y., Qin, B., Zhang, Y., Wang, W., Woolway, R. I., Piao, S., & Jeppesen, E. (2023). Climate Change Drives Rapid Warming and Increasing Heatwaves of Lakes. Science Bulletin (Beijing), 68(14), 1574–1584. https://doi.org/10.1016/j.scib.2023.06.028

Warrack, S., Challis, J. K., Hanson, M. L., & Rennie, M. D. (2017). Microplastics Flowing Into Lake Winnipeg: Densities, Sources, Flux, and Fish Exposures. Proceedings of Manitoba's Undergraduate Science and Engineering Research, 3.

Wassenaar, L. I., & Rao, Y. R. (2012). Lake Winnipeg: the Forgotten Great Lake. Journal of Great Lakes Research, 38, 1-5. https://doi.org/10.1016/j.jglr.2012.04.004

Waters, C. N., Zalasiewicz, J., Summerhayes, C., Barnosky, A. D., Poirier, C., Gałuszka, A., ... & Wolfe, A. P. (2016). the Anthropocene is Functionally and Stratigraphically Distinct from the Holocene. Science, 351(6269), aad2622.

Waters, S., Webster-Brown, J. G., & Hawes, I. (2021). the Release of Legacy Phosphorus from Deforestation-Derived Sediments in Shallow, Coastal Lake Forsyth/Te Roto o Wairewa. New Zealand Journal of Marine and Freshwater Research, 55(3), 446-465. https://doi.org/10.1080/00288330.2020.1804408

Wehr, J.D.,Sheath, R.G., Kociolek, J.P. (2015) Freshwater Algae of North America: Ecology and Classification. Elsevier Inc.

Weinke, A. D., & Biddanda, B. A. (2019). Influence of Episodic Wind Events on Thermal Stratification and Bottom Water Hypoxia in a Great Lakes Estuary. Journal of Great Lakes Research, 45(6), 1103–1112. https://doi.org/10.1016/j.jglr.2019.09.025

Wells, M. L., Trainer, V. L., Smayda, T. J., Karlson, B. S. O., Trick, C. G., Kudela, R. M., … Cochlan, W. P. (2015). Harmful Algal Blooms and Climate Change: Learning from the Past and Present to Forecast the Future. Harmful Algae, 49, 68–93. https://doi.org/10.1016/j.hal.2015.07.009

Werther, M., Odermatt, D., Simis, S. G. H., Gurlin, D., Jorge, D. S. F., Loisel, H., Hunter, P. D., Tyler, A. N., & Spyrakos, E. (2022). Characterising Retrieval Uncertainty of

chlorophyll-a Algorithms in Oligotrophic and Mesotrophic Lakes and Reservoirs. ISPRS Journal of Photogrammetry and Remote Sensing, 190, 279–300. https://doi.org/10.1016/j.isprsjprs.2022.06.015

Weyhenmeyer, G. A., Westöö, A. K., & Willén, E. (2008). Increasingly Ice-Free Winters and Their Effects on Water Quality in Sweden's Largest Lakes. European Large Lakes Ecosystem Changes and Their Ecological and Socioeconomic Impacts, 111-118. https://doi.org/10.1007/s10750-007-9188-9

Whipple, G.C. (1899). the Microscopy of Drinking Water. New York: Wiley.

Wilken, S., Soares, M., Urrutia-Cordero, P., Ratcovich, J., Ekvall, M. K., Van Donk, E., & Hansson, L. A. (2018). Primary Producers or Consumers? Increasing Phytoplankton Bacterivory Along a Gradient of Lake Warming and Browning. Limnology and Oceanography, 63, S142–S155. https://doi.org/10.1002/lno.10728

Williams, F. M. (1967). A Model of Cell Growth Dynamics. Journal of Theoretical Biology, 15(2), 190-207.

Williams, M. R., & King, K. W. (2020). Changing Rainfall Patterns Over the Western Lake Erie Basin (1975–2017): Effects on Tributary Discharge and Phosphorus Load. Water Resources Research, 56(3). https://doi.org/10.1029/2019WR025985

Wiltsie, D., Schnetzer, A., Green, J., Borgh, M. Vander, & Fensin, E. (2018). Algal Blooms and Cyanotoxins in Jordan Lake, North Carolina. Toxins, 10(2). https://doi.org/10.3390/toxins10020092

Winter, J. G., Desellas, A. M., Fletcher, R., Heintsch, L., Morley, A., Nakamoto, L., & Utsumi, K. (2011). Algal Blooms in Ontario, Canada: Increases in Reports Since 1994. Lake and Reservoir Management, 27(2), 105–112. https://doi.org/10.1080/07438141.2011.557765

Woo, M. & Rouse, W. R. (2008). MAGS Contribution to Hydrologic and Surface Process Research. in Cold Region Atmospheric and Hydrologic Studies. The Mackenzie GEWEX Experience (Vol. 2, pp. 9–38). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75136-6_2

Woolway, R. I., & Merchant, C. J. (2018). Intralake Heterogeneity of Thermal Responses to Climate Change: a Study of Large Northern Hemisphere Lakes. Journal of Geophysical Research: Atmospheres, 123(6), 3087–3098. https://doi.org/10.1002/2017JD027661

Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., & Sharma, S. (2020). Global Lake Responses to Climate Change. Nature Reviews Earth & Environment, 1(8), 388-403. https://doi.org/10.1038/s43017-020-0067-5

Woolway, R. I., Sharma, S., Weyhenmeyer, G. A., Debolskiy, A., Golub, M., Mercado-Bettín, D., ... & Jennings, E. (2021). Phenological Shifts in Lake Stratification Under Climate Change. Nature Communications, 12(1), 2318. https://doi.org/10.1038/s41467-021-22657-4

Woolway, R. I., Weyhenmeyer, G. A., Schmid, M., Dokulil, M. T., de Eyto, E., Maberly, S. C. … Merchant, C. J. (2019). Substantial Increase in Minimum Lake Surface Temperatures Under Climate Change. Climatic Change, 155(1), 81–94. https://doi.org/10.1007/s10584-019-02465-y

Woolway, R. L., & Merchant, C. J. (2019). Worldwide Alteration of Lake Mixing Regimes in Response to Climate Change. Nature Geoscience, 12(4), 271–276. https://doi.org/10.1038/s41561-019-0322-x

Woolway, R. L., Meinson, P., Nõges, P., Jones, I. D., & Laas, A. (2017). Atmospheric Stilling Leads to Prolonged Thermal Stratification in a Large Shallow Polymictic Lake. Climatic Change, 141(4), 759–773. https://doi.org/10.1007/s10584-017-1909-0

Woolway, S. S., Sharma, S., & Smol, J. P. (2022). Lakes in Hot Water: the Impacts of a Changing Climate on Aquatic Ecosystems. Bioscience, 72(11), 1050–1061. https://doi.org/10.1093/biosci/biac052

Worth, D. (1989). AVHRR Satellite Monitoring of Algal Blooms and Turbidity on Lake Okeechobee, Florida. Technical Publication #89-2. Environmental Sciences Division, Resource Planning Department, South Florida Water Management District, West Palm Beach, Florida.

Wu, M., Zhang, W., Wang, X., & Luo, D. (2009). Application of MODIS Satellite Data in Monitoring Water Quality Parameters of Chaohu Lake in China. Environmental Monitoring and Assessment, 148(1–4), 255–264. https://doi.org/10.1007/s10661-008-0156-2

Wu, Y., C.R. Duguay, and L. Xu, 2021. Assessment of Machine Learning Classifiers for Global Lake Ice Cover Mapping from MODIS TOA Reflectance Data. Remote Sensing of Environment, 253, 112206, https://doi.org/10.1016/j.rse.2020.112206.

Xi, H., Losa, S. N., Mangin, A., Soppa, M. A., Garnesson, P., Demaria, J., Liu, Y., d'Andon, O. H. F., & Bracher, A. (2020). Global Retrieval of Phytoplankton Functional Types Based on Empirical Orthogonal Functions Using CMEMS Globcolour Merged Products and Further Extension to OLCI Data. Remote Sensing of Environment, 240, 111704–. https://doi.org/10.1016/j.rse.2020.111704

Xie, C., Zhang, X., Zhuang, L., Zhu, R., & Guo, J. (2022). Analysis of Surface Temperature Variation of Lakes in China Using MODIS Land Surface Temperature Data. Scientific Reports, 12(1), 2415. https://doi.org/10.1038/s41598-022-06363-9

Xie, X., Tian, Y., & Wei, G. (2023). Deduction of Sudden Rainstorm Scenarios: Integrating Decision Makers' Emotions, Dynamic Bayesian Network and DS Evidence Theory. Natural Hazards (Dordrecht), 116(3), 2935–2955. https://doi.org/10.1007/s11069-022-05792-z

Xu, J., Pan, J., & Devlin, A. T. (2023). Variations in chlorophyll-a Concentration in Response to Hydrodynamics in a Flow-Through Lake: Remote Sensing and Modeling

Studies. Ecological Indicators, 148, 110128.
https://doi.org/10.1016/j.ecolind.2023.110128

Xu, S., Li, S., Tao, Z., Song, K., Wen, Z., Li, Y., & Chen, F. (2022). Remote Sensing of chlorophyll-a in Xinkai Lake Using Machine Learning and GF-6 WFV Images. Remote Sensing, 14(20), 5136. https://doi.org/10.3390/rs14205136

Xuan Vinh, N., Chetty, M., Coppel, R., & Wangikar, P. P. (2012). Gene Regulatory Network Modeling Via Global Optimization of High-Order Dynamic Bayesian Network. BMC bioinformatics, 13, 1-16. https://doi.org/10.1186/1471-2105-13-131

Xue, K., Ma, R., Duan, H., Shen, M., Boss, E., & Cao, Z. (2019). Inversion of Inherent Optical Properties in Optically Complex Waters Using Sentinel-3A/OLCI Images: a Case Study Using China's Three Largest Freshwater Lakes. Remote sensing of environment, 225, 328-346. https://doi.org/10.1016/j.rse.2019.03.006

Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., & Chen, F. (2022). A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges. Remote Sensing, 14(8), 1770. https://doi.org/10.3390/rs14081770

Yang, K., Yu, Z., Luo, Y., Zhou, X., & Shang, C. (2019). Spatial-Temporal Variation of Lake Surface Water Temperature and Its Driving Factors in Yunnan-Guizhou Plateau. Water Resources Research, 55(6), 4688–4703. https://doi.org/10.1029/2019WR025316

Yang, X., Chen, J., & Yu, Y. (2023). Analysis of the Bidirectional Characteristic of Radiation of Flat and Rough Water–Air Interfaces Based on the Theory of Radiative Transfer. Sustainability, 15(1), 140. https://doi.org/10.3390/su15010140

Yao, X., Li, L., Zhao, J., Sun, M., Li, J., Gong, P., & An, L. (2015). Spatial-Temporal Variations of Lake Ice in the Hoh Xil Region from 2000 to 2011. Dili Xuebao/Acta Geographica Sinica, 70(7), 1114–1124. https://doi.org/10.11821/dlxb201507008

Zeebe, R., & Wolf-Galdrow, D. (2001). CO2in Seawater: Equilbrium, Kinetics, Isotopes. Elsevier Oceanography Series (Vol. 65).

Zeng, C., & Binding, C. (2019). The Effect of Mineral Sediments on Satellite chlorophyll-a Retrievals from Line-Height Algorithms Using Red and Near-Infrared Bands. Remote Sensing (Basel, Switzerland), 11(19), 2306–. https://doi.org/10.3390/rs11192306

Zhai, S. Y., Huang, P., Marshall, J. C., Lobegeiger, J., Cramp, R. L., Parisi, M. A., Franklin, C. E., Prior, A., Kurucz, K., & Hipsey, M. R. (2023). Modelling Prolonged Stratification and Hypoxia in Dryland River Waterholes During Drought Conditions. Inland Waters (Print), 1–21. https://doi.org/10.1080/20442041.2023.2213629

Zhan, X., Bo, Y., Zhou, F., Liu, X., Paerl, H. W., Shen, J., Wang, R., Li, F., Tao, S., Dong, Y., & Tang, X. (2017). Evidence for the Importance of Atmospheric Nitrogen Deposition to Eutrophic Lake Dianchi, China. Environmental Science & Technology, 51(12), 6699–6708. https://doi.org/10.1021/acs.est.6b06135

Zhang, Q. Q., Xu, Y. P., Tian, Y., & Zhang, X. J. (2012). Risk-Based Water Quality Decision-Making Under Small Data Using Bayesian Network. Journal of Central South University, 19(11), 3215-3224. https://doi.org/10.1007/s11771-012-1398-2

Zhang, R., Luo, L., Pan, M., He, F., Luo, C., Meng, D., Li, H., Li, J., Gong, F., Wu, G., Chen, L., Zhang, J., & Sun, T. (2022). Estimations of Water Volume and External Loading Based on DYRESM Hydrodynamic Model at Lake Dianchi. Water (Basel), 14(18), 2832–. https://doi.org/10.3390/w14182832

Zhang, S., & Pavelsky, T. M. (2019). Remote Sensing of Lake Ice Phenology Across a Range of Lakes Sizes, ME, USA. Remote Sensing, 11(14), 1718.

Zhang, X., Flato, G., Kirchmeier-Young, M., Vincent, L., Wan, H., Wang, X., Rong, R., Fyfe, J., Li, G., Kharin, V.V. (2019a). Changes in Temperature and Precipitation Across Canada; Chapter 4 in Bush, E. and Lemmen, D.S. (Eds.) Canada's Changing Climate Report. Government of Canada, Ottawa, Ontario, pp 112-193.

Zhang, X., Li, B., Xu, H., Wells, M., Tefsen, B., & Qin, B. (2019b). Effect of Micronutrients on Algae in Different Regions of Taihu, a Large, Spatially Diverse, Hypereutrophic Lake. Water Research, 151, 500–514. https://doi.org/10.1016/j.watres.2018.12.023

Zhang, Y., Lin, S., Qian, X., Wang, Q. G., Qian, Y., Liu, J., & Ge, Y. (2011). Temporal and Spatial Variability of chlorophyll-a Concentration in Lake Taihu Using MODIS Time-Series Data. Hydrobiologia, 661, 235-250. https://doi.org/10.1007/s10750-010-0528-9

Zhang, Y., Ma, R., Duan, H., Loiselle, S., Zhang, M., & Xu, J. (2016). a Novel MODIS Algorithm to Estimate chlorophyll-a Concentration in Eutrophic Turbid Lakes. Ecological Indicators, 69, 138–151. https://doi.org/10.1016/j.ecolind.2016.04.020

Zhao, C., Brissette, F., Chen, J., & Martel, J. L. (2020). Frequency Change of Future Extreme Summer Meteorological and Hydrological Droughts Over North America. Journal of Hydrology, 584, 124316. https://doi.org/10.1016/j.jhydrol.2019.124316

Zhao, D., Feng, L., & He, X. (2023). Global Gridded Aerosol Models Established for Atmospheric Correction Over Inland and Nearshore Coastal Waters. Journal of Geophysical Research: Atmospheres, 128(16). https://doi.org/10.1029/2023JD038815

Zhou, J., Han, X., Qin, B., & Zhu, G. (2021). Responses of Alkaline Phosphatase Activity to Wind-Driven Waves in a Large, Shallow Lake: Implications for Phosphorus Availability and Algal Blooms. Journal of Environmental Sciences, 99, 143–150. https://doi.org/10.1016/j.jes.2020.06.022

Zhu, L., Suomalainen, J., Liu, J., Hyyppä, J., Kaartinen, H., & Haggren, H. (2018). A Review: Remote Sensing Sensors. Multi-Purposeful Application of Geospatial Data, 19-42.

Zhu, W.D., Qian, C.Y., He, N.Y., Kong, Y.X., Zou, Z.Y., & Li, Y.W. (2022). Research on chlorophyll-a Concentration Retrieval Based on BP Neural Network Model—Case

Study of Dianshan Lake, China. Sustainability, 14(14), 8894–.
https://doi.org/10.3390/su14148894

Ziyad, J., Goïta, K., Magagi, R., Blarel, F., & Frappart, F. (2020). Improving the Estimation
of Water Level Over Freshwater Ice Cover Using Altimetry Satellite Active and
Passive Observations. Remote Sensing, 12(6), 967–.
https://doi.org/10.3390/rs12060967

Appendices

Appendix A: In Depth Review of Machine Learning and Bio-Optical Techniques

A1. Artificial Neural Networks (ANN)

One of the most commonly implemented methods for the retrieval of algal biomass is the ANN method. The ANN is designed to mimic the way the human brain processes data, first established by Dr. Frank Rosenblatt in the 1950s. The inputs of the ANN are connected to nodes, where the patterns of node activation derive the classification of the output, and as such may derive non-linear and non-deterministic relationships (Mani & Srinivasan, 1997). The structure of the network is composed of the input, hidden and output layer, with the hidden layer often described as a "black box" (Cao et al., 2022). This is because the hidden layer represents a complex set of equations to derive the activation of each node and then to determine the pattern of the activation (Chang & Bai, 2018). The number of hidden layers and nodes, or "neurons" within each hidden layer is decided upon by the user. The higher the number of layers and neurons, the more complex the model and the longer the training time (Chang & Bai, 2018). An ANN with many hidden layers is considered a deep learning method and is commonly used as artificial intelligence for classification methods (*e.g.*, facial recognition), and requires a significant amount of training samples (Chang & Bai, 2018). For the purpose of algal biomass retrieval, typically only 1-2 hidden layers are required, as too many will result in over fitting of the model. However, the more training samples, the better the better the model, as the architecture of the ANN has a high tolerance for error and excels at identifying patterns in big data, using either supervised or unsupervised training (Chang et al., 2019). This is particularly beneficial for phytoplankton retrieval, as the change in IOPs can cause significant errors in which traditional regression methods are not capable of resolving patterns. The non-linear nature of freshwater optics is well suited to adaptive learning models such as the ANN. Typical chl-a/PC retrieval models will utilize a single layer feedforward structure design.

244

The structure of the single layer feedforward neural network (SLFN) utilizes a training set of x and y data ($\{x_i, y_i\}$) and is defined by Chang et al. (2019) as the following equation:

$$\sum_{i=1}^{L} v_i g_i(x_j) = \sum_{i=1}^{L} v_i g(a_i \cdot x_j + b_i) \ (j = 1,2,3, \dots, N) \tag{A1.1}$$

where L represents the number of hidden nodes, j the number of input samples, $a_i = [a_{i1}, a_{i2}, \dots, a_{iN}]^T (i = 1,2,3, \dots L)$ represents the weights of the input to the ith hidden node vectors, while $v_i = [v_{i1}, v_{i2}, \dots, v_{iN}]^T (i = 1,2,3, \dots L)$ represents the weights of the ith hidden node vectors to the output, $g_i(x_j)$ represents the activation function and b_i represents the bias of each hidden node vector. The activation function defines the rate of activation expressed by each node given the weight and bias. Therefore, each node is defined as:

$$Y = \sum (weight * input) + bias \tag{A1.2}$$

while the activation function can be expressed by various functions such as:

- cosine: $g_i(x_j) = \frac{1}{1+e^{-x}}$
- tanh: $g_i(x_j) = \frac{2}{1+e^{-2x}} - 1$
- linear: $g_i(x_j) = cx$
- piece-wise (*i.e.* rectified linear activation function (ReLU)): $g_i(x_j) = \max(0, x)$ etc.,

in which each have their own benefits given the purpose of the SLFN and the input data. The activation function allows for the calculation of an output (Y) for the number of training inputs (N) and is expressed as:

$$HV = Y \tag{A1.3}$$

where H represents the hidden layer output matrix and V is the predicted value. The dimensions of H is expressed as the number of training samples (N) and the number of hidden nodes (L) such as:

$$H = \begin{bmatrix} g(a_1 \cdot x_1 + b_1) & \cdots & g(a_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(a_1 \cdot x_N + b_1) & \cdots & g(a_L \cdot x_N + b_L) \end{bmatrix} \tag{A1.4}$$

$$V = \begin{bmatrix} V_1^T \\ \vdots \\ V_L^T \end{bmatrix}_{L \times m}, \qquad Y = \begin{bmatrix} Y_1^T \\ \vdots \\ Y_N^T \end{bmatrix}_{N \times m} \tag{A1.5}$$

This process then seeks to minimize the difference between the approximated values and the true values, optimizing the model (or optimization function). Typically, this is done using stochastic gradient descent or variants such as Root Mean Square Propagation (RMSProp). Gradient descent looks to find at which point the model has the lowest possible error, where the weights are adjusted toward a lower error. For example, RMSProp can be defined by the following equation:

$$v_t = \rho v_{t-1} + (1 - \rho) \cdot g_t^2 \tag{A1.6}$$

$$\Delta \omega_t = - \frac{\eta}{\sqrt{v_t + \epsilon}} \cdot g_t \tag{A1.7}$$

$$\omega_{t+1} = \omega_t + \Delta \omega_t \tag{A1.8}$$

where v_t is the exponential mean of squares of each gradient, ρ is the momentum (user defined), g_t is the sum of squares of each gradient, ω is the weight of each input, t is the time step, η is the learning rate (user defined), and ϵ is the error term. A random set of weights is set as an initial placement, where the optimization function then calculates the multivariate direction where lower error occurs η and ρ are user defined variables which determine how large each step is in this process, and how much momentum the step has. Momentum would allow the step to continue moving even if the error were to increase. Momentum is important, as the purpose of gradient descent is to find the global minimum error, to which the model then has optimal fit. However, the objective function may instead find a local minimum and assume that it is the best fit for the model, which may not be the case. Momentum helps in finding the global minimum. Additionally, the process is repeated across multiple attempts known as epochs, which is also user defined. The epoch with the lowest error is determined as the optimal

number of training batches the SLFN must undergo. While the increase in the number of epochs is useful in determining the global minimum, it does increase training time, and if too many batches are used, it can result in overfitting. Training time can be reduced using mini-batches over all the training data with each step. Overfitting may occur when the total loss (measured in a user-defined error metric) of internal validation either exceeds that of the training error, or the difference is too great, indicating that the application of the model to new data would result in greater error than the model itself (which can be tested using a testing dataset). The use of too many nodes and hidden layers may also result in overtraining of the model (Rao et al., 2020). The number of recommended layers depends on the complexity of the problem and the number of training samples. A neural network model with a significant number of hidden layers and nodes is instead classified as a deep learning model. While not typically used in the retrieval of phytoplankton biomass, certain deep learning models such as a convolution neural network (CNN) are used.

The process of feed forward artificial neural networks uses either single pixels, or a pixel window mean of the sampling location as the input vector. However, the CNN accounts for both the spatial and temporal patterns surrounding each input variable. A traditional CNN is used as a process of classification, in which all images are convoluted and then pooled to a smaller matrix. The process is repeated until all pixels are presented in a fully connected layer (1:n matrix). Convolution is the process of applying a moving window kernel of predetermined size features to calculate a new pixel value, compressing the size of the image to the convolved feature. Multiple kernels may be applied to train the model to detect specific features. Pooling (either max or mean) is also used to reduce the feature size, increase computational speed, and reduce noise. The max pooling method will return a maximum value within the pooling window (size determined by the user), while the mean returns the average. The pooling window will move across the convolved feature, with a predetermined stride value. The fully connected layer will then act as an input to the neural network, where the process is the same as described for the SLFN. For regression analysis, the CNN is as defined by Dong et al. (2020) and Li et al. (2018):

$$map_{l,j}^{x,y} = f \left(\sum_{m} \sum_{h=o}^{H_{l-1}} \sum_{w=o}^{W_{l-1}} k_{l,j,m}^{h,w} map_{(l-1),m}^{(x+h)(y+w)} + b_{l,j} \right) \qquad (A1.9)$$

where $map_{l,j}^{x,y}$ is the convolved output of each x and y pair per the number of node layers (l) and total number of samples (j), $f(x)$ is the activation function, m is the number of convolved feature maps, h and w are the positions of the convolution kernel, $k_{l,j,m}^{h,w}$ is the kernel value at the h and w position which is then connected to $map_{(l-1),m}^{(x+h)(y+w)}$, which is the mth convolved feature map for each node ($l-1$), H and W are the height and width of the kernel and $b_{l,j}$ is the bias at each node to input connection. The pooling layer would then use the following equation as described by Pyo et al. (2019):

$$f_m(x^1) = \max(x_i^l) \qquad (A1.10)$$

where $f_m(x^1)$ is the max pooling result for node layer of the CNN. The difference between a CNN for regression analysis to derive algal biomass and that of image classification is that the regression CNN as described by Dong et al. (2020) does not use the entire image for training, and instead only uses an initial 7x7 matrix surrounding each input in situ observation, which is then convolved down to the fully connected layer. Too large a window surrounding each input would potentially increase noise within the model, as pixels further away from the target location have less impact on the target location. Various studies have found varying CNN performance results with testing data r^2 of 0.35 (Syariz et al., 2019), 0.98 (Dong et al., 2020), and 0.73 (Pyo et al., 2019). Compared to the bio-optical models of Simis et al. (2005), Duan et al. (2012) and Pyo et al. (2018), CNN models have had greater simulation accuracy and between 12-62% lower error (Pyo et al., 2019). In comparison to other machine learning models, Dong et al. (2020) had found much higher performance of the CNN than ANN, RF and SVR, with between 30-60% lower errors. However, CNN is found to have high variance in the results depending on the training data (Dong et al., 2020, Syariz et al., 2019).

A2. Random Forest (RF)

The RF method utilizes a series of regression trees, which separates the inputs into clusters based on the total error observed within each cluster. For application in phytoplankton biomass retrieval by remote sensors, multivariate band inputs may be used as predictor variables $(x_1, x_2 \ldots x_n)$ to the response variable of chl-a or PC (y). Regression trees as described by Breimen et al. (1984) use a series of linear regressions to determine predictor thresholds which produce the lowest possible error for the response variable. The predictor variable split points $(t_1, t_2 \ldots t_j)$ are binary and result in terminal nodes of the response variable $(Y_1, Y_2 \ldots Y_L)$, where the binary split only stops due to a predefined criterion (*e.g.*, minimum number of observations). Regression trees tend to overfit the training data, have low bias, and high variance, whereas other empirical measurement options are preferred. However, RF methods use many regression trees, in which the input training observations are bagged, in which each tree consists of different training data. The process of bagging helps reduce variance in the testing data and shows improved results compared to a single regression tree. Breiman (1996) describes the bagging process where the total data set $(\mathcal{L})$ consists of observation pairs $\{(x_n, y_n), n = 1, \ldots N\}$, where $y = \varphi(x, \mathcal{L})$ (in which $\varphi(x, \mathcal{L}_k)$ is a predictor) and is derived from multiple bags $(\{\mathcal{L}_k\})$ (in which k is the number of bags). As this method is used as a regressor, y is predicted using the average of $\varphi(x, \mathcal{L}_k)$ as described by Breiman (1996) in the following equation:

$$\varphi_A(x) = E_{\mathcal{L}}\varphi(x, \mathcal{L}) \tag{A2.1}$$

where A is the aggregation of the predictor (x) and $E_{\mathcal{L}}$ is the expectation over $\mathcal{L}$, in which there are multiple learning sets of $\mathcal{L}$. As there is often only one given set of data, an imitation of multiple learning sets is to bootstrap aggregate the data known as bagging and is given as the following equation:

$$\varphi_B(x) = av_B\varphi(x, \mathcal{L}^{(B)}) \tag{A2.2}$$

where $\mathcal{L}^{(B)}$ is a repeated bootstrap of the input data with replacements of a total of N observations. In this case of bagging, the term replacement indicates that a randomized

observation pair $\{(x_n, y_n), n = 1, \dots N\}$ may either not be used at all in the model or used more than once across all bags and even within the same bag (Breiman, 1996). Each terminal node or leaf represents the average of $\varphi(x, \mathcal{L}_k)$, which aides in reducing variance of the prediction results (Hastie et al., 2001). The RF is a collection of regression trees, in which each tree represents the subset of the data $\mathcal{L}^{(B)}$. A single tree predictor is defined as $\varphi(x, \mathcal{L}_k)$, where the random forest will take the average over k (Brieman, 2001). The mean-square generalization error of each output is defined as:

$$E_{x,y}(y - \varphi(x))^2 \tag{A2.3}$$

where y is the true value, and $E_{x,y}$ is the error term for each observation pair. Therefore, as the number of bags or trees (k) becomes very large (assumed infinite), the RF is defined by the Law of Large Numbers:

$$E_{x,y}(y - av_B\varphi(x, \mathcal{L}^{(B)}))^2 \rightarrow E_{x,y}(y - E_{\mathcal{L}}\varphi(x, \mathcal{L}))^2 \tag{A2.4}$$

where the error observed in the trees represents the error of the data set. This can be summarized as the following:

$$PE^*(tree) = E_{\mathcal{L}}E_{x,y}(y - \varphi(x, \mathcal{L}))^2 \tag{A2.5}$$

where PE^* is the estimated generalized error of each tree and $E_{\mathcal{L}}$ is the error of the dataset. For the generalization error of the forest is summarized as:

$$PE^*(Forest) = E_{x,y}(y - E_{\mathcal{L}}\varphi(x, \mathcal{L}))^2 \tag{A2.6}$$

If it is assumed that for all of $\mathcal{L}$ the RF model is unbiased, in which the error of Y is equal to that of the predictor ($\because EY = E_x\varphi(X, \mathcal{L})$) then the generalized error of the forest is less than or equal to the weighted correlation ($\bar{\rho}$) of the residuals of the data ($Y - \varphi(x, \mathcal{L})$) and bagged data ($Y - \varphi(x, \mathcal{L}^B)$):

$$PE^*(Forest) \leq \bar{\rho}PE^*(tree) \tag{A2.7}$$

where a good RF model will show low correlation between the residuals and tree error (Breiman, 2001). The randomization due to bagging helps in reducing the correlation of the residuals and error (Segal, 2003).

The RF model may also be used as a method of classification and has been shown to be effective in defining optical water types (Pereira-Sandoval et al., 2022) and algal bloom identification (Hill et al., 2020). Similar to the regressor, RF classification is constructed using a series of binary trees, which are composed of binary partitions called nonterminal nodes, whereby each predictor variable (X_i) is split into two descendent nodes by the set of defined categories $(S_i = \{S_{i,1}, \dots, S_{i,m}\})$ of the predicted variable (S) (Cutler et al., 2011). The left descendent node contains a subset of the categories, the remainder to the right node, where each split is defined by the following equation:

$$X_i \in S \subset S_i \tag{A2.8}$$

To determine the predictor split, a criterion is often used to determine the best partition, and is often determined using a GINI index as described in the following equation:

$$Q = \sum_{k \neq k'}^{K} \hat{P}_k \hat{P}_{k'} \tag{A2.9}$$

Where $\hat{P}_k$ = proportion of observations of each class (k) within the descendent node, as described in the following equation:

$$\hat{P}_k = \frac{1}{n} \sum_{i=1}^{n} I(y_i = k) \tag{A2.10}$$

Where the larger the derived value, the lower the purity of the node, therefore, the split with the lowest derived value is selected. This procedure reoccurs until the stopping criteria are met, such as a predefined number of terminal nodes. The accuracy of the tree is assessed by determining the number of observations placed in terminal nodes which match the designated output class (Cutler et al., 2011).

A3. Bio-Optical Models

Establishing a forward model would allow water quality parameters to be predicted without the use of *in situ* observation data. Common forward models include radiative transfer equations and reflectance approximation, where simulated data derive coefficients to predict

remote sensing reflectances (Matthew, 2011). A forward model in freshwater systems is difficult to derive due to the spatial/temporal change and variability in IOPs, in which accurate profiles have not currently been developed. The semi-analytical approach utilizes both empirical and analytical measurements. An inversion model of AOPs, the relationship to IOPs along with remote sensing measurements, is necessary to construct a forward model (Ogashawara et al., 2017). Bio-optical models are predominantly a semi-analytical method in which remotely sensed measurements estimate the IOPs and $b_b(\lambda)$ of water. The inversion of IOPs utilizes remotely sensed measurements to derive an absorption coefficient at a given wavelength ($a_i(\lambda)$) of water ($a_w(\lambda)$), chl-a ($a_{Chl}(\lambda)$) and/or PC ($a_{PC}(\lambda)$) (Simis et al., 2005; Zhou et al., 2019). The semi-analytical approach has been found to have higher inversion precision compared to that of empirical regression methods (Allan et al., 2015; Kong et al., 2015). The application of $a(\lambda)$ and $b_b(\lambda)$ is based on the Kubelka-Munk remission function in which the change in turbidity of a medium is proportionate to the ratio of $a(\lambda)$ and $b_b(\lambda)$ (Ogashawara et al., 2017). Of the bio-optical models developed overtime, only case-I waters have developed a forward model (analytical approach), while case-II waters required an inversion to derive IOPs (semi-analytical) (Morel & Prieur, 1977). The analytical bio-optical model requires a predetermined spectral shape function of each parameter along with an estimation of total $b_b(\lambda)$. The relationship between the AOPs and IOPs was first defined by Gordon et al. (1975) and is defined as:

$$R(0^-, \lambda) = f b_b / (a + b_b) \qquad (A3.1)$$

where $R(0^-, \lambda)$ is the subsurface irradiance of a given wavelength at a depth of 0 and f is a light field factor (Gorden et al., 1975; Simis et al., 2005). Based on the definition of AOP by Gordon, Morel and Prieur (1977) attempted to define the coefficients of a and b_b in case-I and case-II waters. Morel and Prieur defined scattering ($b(\lambda)$) as the sum of the scattering of water ($b_w(\lambda)$) and particles ($b_p(\lambda)$):

$$b(\lambda) = b_w(\lambda) + b_p(\lambda) \qquad (A3.2)$$

Based on this relationship, it was then assumed that $b_b(\lambda)$ was similarly equal to the sum of the backscattering of water $(b_{bw}(\lambda))$ and the backscattering of particles $(b_{bp}(\lambda))$:

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda) \tag{A3.3}$$

in which the scattering due to particles is a factor of particle size. Based on the understanding of how $R(0^-, \lambda)$ is a factor of a and b_b, and that b_b is a factor of $b_{bw}(\lambda)$ and $b_{bp}(\lambda)$, a formulae to model the $R(0^-, \lambda)$ is given for case-I waters:

$$R(\lambda) = \frac{b_{bw}(\lambda) + b_{bp}(\lambda)}{a_w(\lambda)} \tag{A3.4}$$

Given the assumption of case-I water conditions, the modelled $R(0^-, \lambda)$ visible spectra (400-700nm) indicates high reflectance at shorter wavelengths and low reflectance at longer wavelengths due to $a_w(\lambda)$. As $b_w(\lambda)$ cannot change, only $b_{bp}(\lambda)$ can change, in which higher concentrations create a greater rate of change in $R(0^-, \lambda)$ visible spectra at longer wavelengths compared to shorter ones. The modelled spectra closely resembled the observed spectra of the open ocean. To model $R(0^-, \lambda)$ for turbid waters, the spectra is additionally a result of the absorption of phytoplankton $(a_{phyt}(\lambda))$ which is proportional to chl-a, and absorption of particles not proportional to chl-a $(a_p(\lambda))$, to formulate the case-II formulae:

$$R(\lambda) = \frac{b_w(\lambda) + b_{bp}(\lambda)}{a_w(\lambda) + a_{phyt}(\lambda) + a_p(\lambda)} \tag{A3.5}$$

The modelled case-II spectra, however, resulted in considerable variance when compared to in situ chl-a measurements, due to the algal cell size, species composition, and composition of optically active NAPs. While a forward model was possible given case-I waters, case-II waters are more complicated. Due to the variance of case-II waters, an inversion model is constructed using observed $R(0^-, \lambda)$ instead, leading to the prevalence of the semi-analytical method.

To predict the IOPs from reflectance, it is assumed that the total amount of absorption is contributed by; $a_w(\lambda)$, $a_{phyt}(\lambda)$, the absorption of tripton (or other NAPs; $a_{NAP}(\lambda)$) and the

absorption of CDOM ($a_{CDOM}(\lambda)$) (Roesler & Perry, 1995). The additive absorption assumption is as described in the following equation:

$$a(\lambda) = a_w(\lambda) + a_{phyt}(\lambda) + a_{NAP}(\lambda) + a_{CDOM}(\lambda) \tag{A3.6}$$

Similarly, it is also then assumed that the $b_b(\lambda)$ is also a product of $b_{bw}(\lambda)$ and $b_{bp}(\lambda)$, as described by Roesler and Perry (1995) in the following equation:

$$b_b(\lambda) = b_w(\lambda) + b_{bp}(\lambda) \tag{A3.7}$$

Each parameter has an absorption rate described within the visible-NIR spectrum. The inversion model is then based on Beer's law, which was applied to bio-optical modelling of chl-a by Smith and Baker (1978). Therefore, the concentration-specific spectral shape must be described, where the increase of the concentration will then increase the IOP. The concentration-specific spectral shape of phytoplankton (*e.g.* $a^*_{phyt}(\lambda)$) can be predicted by the $a_{phyt}(\lambda)$ normalized by the chl-a concentration (Roesler *et al.*, 1989). Therefore, the inverse may predict for the IOPs, in which $a^*_{phyt}(\lambda)$) is multiplied by the concentration of chl-a as described by Roesler et al. (1989) in the following equation:

$$a_{phyt}(\lambda) = chla \times a^*_{phyt}(\lambda) \tag{A3.8}$$

This equation will also equate the absorption of NAP and CDOM as well. Therefore, given a known spectral shape of each parameter, we can expand the equation by Gordon et al. (1975) to the following:

$$R(0^-, \lambda) = \frac{f}{Q} \times \frac{b_{bw}(\lambda) + A_{bp}b^*_{bp}(\lambda)}{b_{bw}(\lambda) + A_{bp}b^*_{bp}(\lambda) + a_w(\lambda) + A_{phyt}a^*_{phyt}(\lambda) + A_{NAP}a^*_{NAP}(\lambda) + A_{CDOM}a^*_{CDOM}} \tag{A3.9}$$

where the absorption specific spectral curves are modelled and the concentrations (A_{bp}, A_{phyt}, A_{NAP}, A_{CDOM}) are to be predicted, Q is the geometric attenuation of light and f is a constant based on the volume scattering function (Sathyendranath et al., 1989; Roesler, 2014). The concentrations then act as an eigenvalue, where the absorption/backscatter specific spectra are the eigenvectors. A regression (either linear or non-linear least square) of the remotely measured spectra and the absorption specific spectral curves is used to derive the eigenvalue.

While hyperspectral sensors can be used to define the optical curves, multispectral sensors use individual bands to estimate a and b_b. Studies using multispectral sensors will also not always include the contribution of CDOM and NAP; either the impacts are considered negligible or correction factors are introduced. The wavelengths used to estimate a and b_b of the bio-optical model varies between studies and sensors. Gons (1999) established a chl-a retrieval bio-optical model using wavelengths measured by a spectroradiometer, and is defined as:

$$[Chl\ a] = \{R(a_w(704) + b_b) - a_w(672) - b_b\}/a^*(672) \qquad \text{(A3.10)}$$

Where R is the reflectance ratio (704/672), b_b is the backscatter estimated at 776nm and a^* is the chl-a specific absorption coefficient. The model returned high inversion accuracy with an r^2 of >0.95 and a standard error of <9 mg m^{-3}. Gons *et al.* (2002) then adapted this model to MERIS and is defined as:

$$[Chl\ a] = \{R_M(a_w(704) + b_b) - a_w(664) - b_b\}/a^*(664) \qquad \text{(A3.11)}$$

where R_M is the reflectance ratio (704/664), and b_b is again estimated at 776nm. The Gons models operate on the assumption that the absorption at 672nm and 664nm is negligible for NAPs and therefore represents the chl-a signal. The Gons et al. (2002) bio-optical model was then adapted by Simis *et al.* (2005) for PC retrieval in which the $a_{Chl}(\lambda)$ and the $a_{PC}(\lambda)$ must be separated. It was first established that both $a_{Chl}(\lambda)$ and $a_{PC}(\lambda)$ signals are equal to the Gons *et al.* (2002) model at the 620nm maximal PC absorption, and is defined as:

$$a_{Chl}(620) + a_{PC}(620)$$

$$= \left(\left\{ \left[\frac{R(709)}{R(620)} \right] \times [a_w(709) + b_b] \right\} - b_b - a_w(620) \right) \times \delta^{-1} \qquad \text{(A3.12)}$$

where δ^{-1} is the correction factor of $a^*(620)$. To isolate $a_{PC}(620)$ from $a_{Chl}(620)$, Simis et al. (2005) added a conversion factor (ε) derived from the absorption difference of chl-a at 620 nm compared to the maximal absorption at 665nm, and is defined as:

$$a_{PC}(620) = \left(\left\{ \left[\frac{R(709)}{R(620)} \right] \times [a_w(709) + b_b] \right\} - b_b - a_w(620) \right) \times \delta^{-1}$$
$$- [\varepsilon \times a_{Chl}(665)]$$

(A3.13)

in which the ratio of $a_{PC}(620)$ and $a^*(620)$ quantifies the PC concentration value. The Simis *et al*. (2005) model also returned a high inversion accuracy with a testing r^2 of 0.94 and a root mean square error (RMSE) of 6.5 mg m^{-3}. Allan *et al*. (2015) provide an alternative bio-optical model developed with Landsat 7 ETM+ imagery, in which the $b_b(\lambda)$ model includes that of water constituents and phytoplankton, which the absorption estimates add CDOM. The Allan *et al.* (2015) $b_b(\lambda)$ model is as follows:

$$b_b(\lambda) = b_{bw}(\lambda) + B_{bp}b^*_p(\lambda)C_p + B_{b\phi}b^*_\phi(\lambda)C_\phi$$

(A3.14)

where $b_{bw}(\lambda)$ is the backscattering coefficient of water, $B_{bp}(\lambda)$ is the backscattering ratio of particulates, $b^*_p(\lambda)$ is the specific coefficient of particulate matter, C_p is the concentration of the particulates, $B_{b\phi}(\lambda)$ is the backscattering ratio of phytoplankton, $b^*_\phi(\lambda)$ is the specific scattering coefficient of phytoplankton and C_ϕ is the concentration of chl-a (Allan et al., 2015). The absorption model is described as:

$$a(\lambda) = a_w(\lambda) + C_\phi a^*_\phi(\lambda) + C_p a^*_p + a_{CDOM}(\lambda)$$

(A3.15)

where $a^*_\phi(\lambda)$ is the specific absorption of phytoplankton, $a^*_p(\lambda)$ is the specific absorption of particulates and $a_{CDOM}(\lambda)$ is the absorption of CDOM at 440nm. The wavelengths at which each parameter was measured were tested in each ETM+ band with an r^2 ranging from 0.56-0.68 with band 3 (660 nm) forming the best model with an RMSE of 10.69 mg m^{-3}. As this model includes a measured concentration of suspended particles and chl-a, there is an emphasis on empirical measurements compared to other bio-optical models. Improved bio-optical models are still ongoing, with the complexities of case-II waters, semi-analytical models which invert the observed reflectance to predict IOPs remain highly accurate compared to other empirical methods. Bio-optical methods, however, require extensive in situ sampling and an understanding of the present lake biogeochemistry, in order to accurately define the absorption

specific spectral curves of each parameter. The complexity and the variability of IOPs in freshwater systems can make the definition of the spectral shapes difficult. Rather than attempting to define the shapes, the "black box" approach of machine learning has recently become a popular option.

Appendix B: Tables and Figures for Chapter 4

Table B.7.1. Annual time-series results for each parameter where β = Theil-Sen slope, se = standard error of a Theil-Sen regression, τ = Mann-Kendall tau value, and p = Mann-Kendall p-value. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
GBL CB	MC	0.005	0.079	0.240	□	LW NB	MC	0.003	1.180	0.018	□
	XC	0.019	0.170	0.380	■		XC	0.006	5.294	0.006	□
	BC	0.063	0.317	0.427	■		BC	-0.060	5.092	-0.029	□
	BD	0.813	10.304	0.203	□		BD	0.020	3.459	0.030	□
	PC	-1.600	20.323	-0.270	□		PC	-0.875	14.518	-0.256	□
	MT	-0.005	0.422	-0.088	□		MT	-0.003	0.502	-0.041	□
	XT	0.041	1.508	0.099	□		XT	0.055	0.968	0.251	□
	PT	-0.333	7.236	-0.174	□		PT	0.200	9.340	0.065	□
	AT	-0.545	7.740	-0.219	□		AT	-0.300	3.076	-0.404	■
	ON	0.000	1.237	0.132	□		ON	-0.455	7.494	-0.196	□
	OF	-0.500	8.783	-0.236	□		OF	0.143	12.827	0.072	□
	DR	-0.600	8.545	-0.232	□		DR	0.833	17.968	0.219	□
GBL SB	MC	0.140	0.569	0.497	■	LW SB	MC	-0.140	1.075	-0.415	■
	XC	0.367	1.334	0.637	■		XC	-0.158	5.557	-0.158	□
	BC	0.142	0.588	0.532	■		BC	0.011	4.233	0.006	□
	BD	0.667	10.191	0.266	□		BD	0.000	2.931	-0.166	□
	PC	-0.767	14.436	-0.153	□		PC	-3.440	55.494	-0.181	□
	MT	-0.022	0.519	-0.170	□		MT	0.018	0.677	0.135	□
	XT	0.075	0.987	0.392	■		XT	0.034	0.872	0.340	■
	PT	0.167	7.837	0.071	□		PT	-0.400	9.005	-0.147	□
	AT	-0.500	7.426	-0.266	□		AT	-0.700	18.279	-0.218	□
	ON	0.000	1.572	0.108	□		ON	-0.385	7.704	-0.155	□
	OF	-0.286	7.583	-0.173	□		OF	0.188	12.658	0.064	□
	DR	-0.462	7.106	-0.218	□		DR	0.400	18.794	0.130	□
GBL AB	MC	0.005	0.082	0.263	□	LW AB	MC	-0.022	1.181	-0.146	□
	XC	0.021	0.175	0.392	■		XC	-0.024	4.334	-0.018	□
	BC	0.070	0.365	0.450	■		BC	-0.014	4.963	-0.006	□
	BD	1.091	15.827	0.248	□		BD	0.000	2.847	-0.030	□
	PC	-1.400	19.786	-0.265	□		PC	-1.250	15.227	-0.260	□
	MT	-0.008	0.433	-0.088	□		MT	0.007	0.486	0.029	□
	XT	0.038	1.497	0.088	□		XT	0.056	0.935	0.246	□
	PT	-0.429	7.960	-0.191	□		PT	0.333	9.304	0.100	□
	AT	-0.444	8.443	-0.161	□		AT	-0.333	5.598	-0.278	□
	ON	0.000	1.393	0.197	□		ON	-0.500	7.445	-0.207	□

Lake	Parameter	β	s$_e$	τ	p	Lake	Parameter	β	s$_e$	τ	p
	OF	-0.500	9.059	-0.244	□		OF	0.333	12.077	0.124	□
	DR	-0.545	8.514	-0.220	□		DR	0.833	18.414	0.202	□
GSL CB	MC	0.066	0.422	0.509	■	LE CB	MC	0.082	0.379	0.626	■
	XC	0.100	0.342	0.673	■		XC	0.249	1.234	0.673	■
	BC	0.223	1.062	0.509	■		BC	0.236	1.093	0.696	■
	BD	-0.643	8.527	-0.314	■		BD	0.000	3.579	0.043	□
	PC	-0.114	12.710	-0.035	□		PC	0.420	62.974	0.088	□
	MT	0.002	0.364	0.018	□		MT	0.015	0.589	0.111	□
	XT	0.064	0.827	0.333	■		XT	0.033	0.819	0.170	□
	PT	0.250	6.547	0.142	□		PT	0.083	14.148	0.047	□
	AT	0.143	8.127	0.089	□		AT	-0.500	29.490	-0.089	□
	ON	-0.083	5.825	-0.060	□		ON	-1.077	16.236	-0.314	■
	OF	-0.300	8.632	-0.195	□		OF	-0.818	26.510	-0.130	□
	DR	-0.200	9.745	-0.136	□		DR	0.000	21.494	0.000	□
GSL EB	MC	0.100	0.690	0.380	■	LE EB	MC	0.039	0.243	0.520	■
	XC	0.229	1.240	0.474	■		XC	0.115	0.651	0.614	■
	BC	0.160	0.742	0.497	■		BC	0.180	1.979	0.450	■
	BD	-0.125	9.373	-0.084	□		BD	-1.406	19.646	-0.317	■
	PC	-1.571	22.077	-0.223	□		PC	3.333	50.784	0.181	□
	MT	-0.015	0.346	-0.193	□		MT	0.013	0.542	0.123	□
	XT	0.055	1.300	0.205	□		XT	0.008	0.786	0.136	□
	PT	0.273	9.620	0.132	□		PT	-0.083	16.271	-0.018	□
	AT	-0.154	5.075	-0.155	□		AT	-0.286	31.042	-0.065	□
	ON	0.000	3.464	-0.024	□		ON	-1.000	23.785	-0.188	□
	OF	-0.222	9.615	-0.059	□		OF	-1.200	35.191	-0.199	□
	DR	0.000	9.375	-0.030	□		DR	-0.333	23.289	-0.096	□
GSL NB	MC	-0.094	1.740	-0.251	□	LE WB	MC	0.066	0.446	0.474	■
	XC	-0.039	3.401	-0.041	□		XC	0.213	1.486	0.567	■
	BC	0.280	1.465	0.450	■		BC	0.259	1.183	0.731	■
	BD	0.700	16.380	0.141	□		BD	0.644	6.603	0.324	■
	PC	-1.867	32.326	-0.171	□		PC	0.571	15.643	0.124	□
	MT	-0.053	0.875	-0.205	□		MT	-0.007	0.524	-0.053	□
	XT	0.070	0.624	0.399	■		XT	0.032	0.907	0.176	□
	PT	0.000	7.897	-0.042	□		PT	-0.333	16.981	-0.088	□
	AT	-0.071	7.288	-0.054	□		AT	-1.769	25.940	-0.228	□
	ON	0.286	11.957	0.101	□		ON	0.333	12.809	0.065	□
	OF	-0.250	7.792	-0.174	□		OF	-0.125	23.980	-0.012	□
	DR	-0.875	11.376	-0.282	■		DR	-0.286	23.870	-0.053	□
GSL AB	MC	0.070	0.392	0.497	■	LE AB	MC	0.078	0.376	0.626	■
	XC	0.106	0.360	0.673	■		XC	0.209	1.227	0.673	■
	BC	0.197	0.937	0.497	■		BC	0.230	1.071	0.684	■
	BD	-0.720	8.322	-0.300	■		BD	0.091	4.181	0.109	□
	PC	-0.440	10.945	-0.249	□		PC	2.875	43.329	0.251	□
	MT	0.001	0.354	0.018	□		MT	0.013	0.582	0.135	□
	XT	0.080	0.863	0.328	■		XT	0.032	0.798	0.194	□
	PT	0.250	5.945	0.143	□		PT	0.000	14.293	0.006	□
	AT	0.000	7.670	0.000	□		AT	-0.500	29.741	-0.059	□
	ON	-0.059	5.047	-0.066	□		ON	-0.556	15.015	-0.124	□
	OF	-0.286	7.815	-0.161	□		OF	-0.455	28.865	-0.118	□
	DR	-0.100	9.212	-0.060	□		DR	-0.143	21.125	-0.060	□
L A ·	MC	0.112	1.239	0.392	■						

258

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	XC	0.242	5.747	0.450	■						
	BC	0.559	4.716	0.509	■						
	BD	0.000	9.292	0.006	□						
	PC	0.400	33.278	0.070	□						
	MT	-0.048	0.634	-0.392	■						
	XT	0.076	1.213	0.275	□						
	PT	0.222	7.700	0.083	□						
	AT	-0.182	9.877	-0.071	□						
	ON	-0.167	6.815	-0.072	□						
	OF	-0.400	9.050	-0.119	□						
	DR	-0.500	12.486	-0.152	□						

Table B.7.2. Spring time-series results for each parameter where β = Theil-Sen slope, s_e = standard error of a Theil-Sen regression, τ = Mann-Kendall tau value, and p = Mann-Kendall p-value. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
GBL CB	MC	0.025	4.079	0.609	■	LW NB	MC	-0.012	1.486	-0.111	□
	XC	0.037	5.488	0.597	■		XC	-0.044	1.946	-0.263	□
	BC	0.748	3.142	0.934	■		BC	-0.300	3.140	-0.520	■
	BD	-0.100	1.018	-0.271	□		BD	-0.214	3.840	-0.234	□
	PC	-1.778	10.139	-0.565	■		PC	0.114	10.280	0.072	□
	MT	-0.011	0.370	-0.117	□		MT	-0.023	0.702	-0.205	□
	XT	-0.005	0.548	-0.023	□		XT	-0.050	2.016	-0.135	□
	PT	0.000	15.422	0.094	□		PT	0.300	10.827	0.297	■
	AT	-0.200	19.429	-0.255	□		AT	0.000	2.722	-0.006	□
	ON	0.000	1.237	0.132	□		ON	-0.455	7.494	-0.196	□
	OF	-0.500	8.783	-0.236	□		OF	0.143	12.827	0.072	□
	DR	-0.600	8.545	-0.232	□		DR	0.833	17.968	0.219	□
GBL SB	MC	0.127	0.886	0.657	■	LW SB	MC	-0.307	3.815	-0.392	■
	XC	0.208	1.797	0.669	■		XC	-0.138	5.203	-0.076	□
	BC	0.374	4.519	0.293	■		BC	-0.274	3.684	-0.205	□
	BD	-0.077	3.704	-0.096	□		BD	0.313	7.030	0.153	□
	PC	-0.111	9.625	-0.154	□		PC	-0.143	7.812	-0.065	□
	MT	0.015	0.734	0.082	□		MT	0.035	1.497	0.123	□
	XT	0.052	1.141	0.164	□		XT	0.089	1.467	0.170	□
	PT	0.000	17.486	0.197	□		PT	0.143	3.660	0.286	□
	AT	0.000	11.879	-0.061	□		AT	0.125	6.344	0.089	□
	ON	0.000	1.572	0.108	□		ON	-0.385	7.704	-0.155	□
	OF	-0.286	7.583	-0.173	□		OF	0.188	12.658	0.064	□
	DR	-0.462	7.106	-0.218	□		DR	0.400	18.794	0.130	□
GBL AB	MC	0.036	3.164	0.282	■	LW AB	MC	-0.010	2.901	-0.041	□
	XC	0.058	4.254	0.422	■		XC	-0.068	3.644	-0.228	□
	BC	0.637	4.636	0.610	■		BC	-0.348	3.280	-0.462	■
	BD	-0.333	2.082	-0.447	■		BD	0.080	4.563	0.084	□
	PC	0.000	12.012	-0.116	□		PC	0.250	9.530	0.136	□
	MT	-0.020	0.454	-0.194	□		MT	-0.036	0.836	-0.193	□

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	XT	-0.007	0.664	-0.124	□		XT	-0.076	1.784	-0.099	□
	PT	0.000	16.597	0.085	□		PT	0.200	3.628	0.256	□
	AT	-0.091	11.836	-0.150	□		AT	0.000	3.985	0.006	□
	ON	0.000	1.393	0.197	□		ON	-0.500	7.445	-0.207	□
	OF	-0.500	9.059	-0.244	□		OF	0.333	12.077	0.124	□
	DR	-0.545	8.514	-0.220	□		DR	0.833	18.414	0.202	□
GSL CB	MC	0.029	0.387	0.298	■	LE CB	MC	0.037	1.117	0.158	□
	XC	0.029	0.331	0.462	■		XC	0.042	2.712	0.123	□
	BC	0.062	1.719	0.135	□		BC	0.033	2.514	0.053	□
	BD	0.000	3.285	-0.073	□		BD	-0.083	6.264	-0.108	□
	PC	-0.314	12.361	-0.243	□		PC	-0.600	9.744	-0.276	□
	MT	0.022	0.428	0.146	□		MT	-0.004	0.766	-0.029	□
	XT	0.108	1.537	0.287	■		XT	-0.003	0.987	-0.041	□
	PT	0.154	17.196	0.297	□		PT	-0.214	2.888	-0.242	□
	AT	-0.118	3.572	-0.229	□		AT	0.083	3.220	0.096	□
	ON	-0.083	5.825	-0.060	□		ON	-1.077	16.236	-0.314	■
	OF	-0.300	8.632	-0.195	□		OF	-0.818	26.510	-0.130	□
	DR	-0.200	9.745	-0.136	□		DR	0.000	21.494	0.000	□
GSL EB	MC	0.168	1.101	0.485	■	LE EB	MC	0.034	0.270	0.392	■
	XC	0.261	1.446	0.556	■		XC	0.051	0.489	0.427	■
	BC	0.178	5.280	0.263	□		BC	0.042	1.678	0.064	□
	BD	0.000	6.819	-0.037	□		BD	0.333	3.307	0.510	■
	PC	2.200	11.814	0.539	■		PC	0.000	11.525	-0.012	□
	MT	-0.015	0.443	-0.135	□		MT	0.015	0.808	0.059	□
	XT	0.001	0.758	0.006	□		XT	-0.019	1.489	-0.059	□
	PT	0.000	25.721	0.060	□		PT	-0.200	3.351	-0.215	□
	AT	-0.200	13.595	-0.220	□		AT	-0.250	4.200	-0.223	□
	ON	0.000	3.464	-0.024	□		ON	-1.000	23.785	-0.188	□
	OF	-0.222	9.615	-0.059	□		OF	-1.200	35.191	-0.199	□
	DR	0.000	9.375	-0.030	□		DR	-0.333	23.289	-0.096	□
GSL NB	MC	-0.007	1.008	-0.006	□	LE WB	MC	0.014	0.527	0.064	□
	XC	0.071	2.380	0.111	□		XC	-0.028	1.774	-0.053	□
	BC	0.225	1.905	0.427	■		BC	0.027	2.284	0.053	□
	BD	0.400	5.424	0.185	□		BD	-0.167	5.746	-0.107	□
	PC	0.273	4.969	0.179	□		PC	0.182	13.394	0.059	□
	MT	-0.049	0.947	-0.088	□		MT	-0.001	0.957	-0.006	□
	XT	0.135	1.531	0.287	■		XT	-0.077	1.490	-0.164	□
	PT	0.125	3.315	0.192	□		PT	0.000	3.670	-0.018	□
	AT	0.500	9.164	0.220	□		AT	0.167	4.528	0.170	□
	ON	0.286	11.957	0.101	□		ON	0.333	12.809	0.065	□
	OF	-0.250	7.792	-0.174	□		OF	-0.125	23.980	-0.012	□
	DR	-0.875	11.376	-0.282	■		DR	-0.286	23.870	-0.053	□
GSL AB	MC	0.034	0.408	0.345	■	LE AB	MC	0.050	0.651	0.216	□
	XC	0.037	0.329	0.439	■		XC	0.030	1.369	0.088	□
	BC	0.077	1.848	0.170	□		BC	0.021	2.425	0.029	□
	BD	-0.017	2.771	-0.098	□		BD	-0.067	5.893	-0.084	□
	PC	-0.167	11.940	-0.220	□		PC	-0.429	9.642	-0.136	□
	MT	0.015	0.466	0.111	□		MT	-0.020	0.763	-0.111	□
	XT	0.088	1.242	0.240	□		XT	-0.006	0.940	-0.053	□
	PT	0.143	16.836	0.219	□		PT	0.000	3.125	-0.030	□
	AT	-0.167	3.691	-0.199	□		AT	0.000	3.227	-0.086	□

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	ON	-0.059	5.047	-0.066	□		ON	-0.556	15.015	-0.124	□
	OF	-0.286	7.815	-0.161	□		OF	-0.455	28.865	-0.118	□
	DR	-0.100	9.212	-0.060	□		DR	-0.143	21.125	-0.060	□
LA AB	MC	0.050	2.291	0.099	□						
	XC	0.096	3.129	0.181	□						
	BC	0.276	5.098	0.170	□						
	BD	-0.250	5.563	-0.254	□						
	PC	0.273	14.369	0.126	□						
	MT	-0.067	2.925	-0.310	■						
	XT	-0.188	2.765	-0.275	□						
	PT	0.167	17.455	0.293	□						
	AT	0.176	4.370	0.158	□						
	ON	-0.167	6.815	-0.072	□						
	OF	-0.400	9.050	-0.119	□						
	DR	-0.500	12.486	-0.152	□						

Table B.7.3. Summer time-series results for each parameter where β = Theil-Sen slope, s_e = standard error of a Theil-Sen regression, τ = Mann-Kendall tau value, and p = Mann-Kendall p-value. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
GBL CB	MC	0.007	0.063	0.333	■	LW NB	MC	0.134	2.603	0.158	□
	XC	0.018	0.117	0.392	■		XC	0.006	10.289	0.006	□
	BC	0.057	0.316	0.427	■		BC	-0.041	10.695	-0.029	□
	BD	-0.364	6.974	-0.183	□		BD	0.286	6.099	0.150	□
	PC	0.000	10.457	0.024	□		PC	0.000	10.889	0.006	□
	MT	0.035	0.981	0.123	□		MT	-0.016	0.716	-0.076	□
	XT	0.077	1.669	0.099	□		XT	0.055	0.960	0.251	□
	PT	-0.444	7.417	-0.256	□		PT	0.200	9.334	0.077	□
	AT	-0.500	8.318	-0.307	■		AT	-0.500	3.022	-0.508	■
	ON	0.000	1.237	0.132	□		ON	-0.455	7.494	-0.196	□
	OF	-0.500	8.783	-0.236	□		OF	0.143	12.827	0.072	□
	DR	-0.600	8.545	-0.232	□		DR	0.833	17.968	0.219	□
GBL SB	MC	0.108	0.507	0.450	■	LW SB	MC	0.087	1.464	0.205	□
	XC	0.298	1.157	0.602	■		XC	0.058	6.440	0.099	□
	BC	0.120	0.507	0.474	■		BC	0.007	8.952	0.006	□
	BD	0.145	6.210	0.101	□		BD	-0.400	11.069	-0.229	□
	PC	0.500	7.805	0.142	□		PC	-0.182	13.301	-0.047	□
	MT	0.006	0.995	0.018	□		MT	-0.008	0.680	-0.053	□
	XT	0.075	1.010	0.392	■		XT	0.034	0.827	0.352	■
	PT	-0.200	8.222	-0.071	□		PT	-0.444	9.745	-0.159	□
	AT	-0.167	6.592	-0.196	□		AT	0.000	11.077	0.000	□
	ON	0.000	1.572	0.108	□		ON	-0.385	7.704	-0.155	□
	OF	-0.286	7.583	-0.173	□		OF	0.188	12.658	0.064	□
	DR	-0.462	7.106	-0.218	□		DR	0.400	18.794	0.130	□
GBL AB	MC	0.007	0.067	0.333	■	LW AB	MC	0.126	2.246	0.146	□
	XC	0.019	0.123	0.380	■		XC	-0.097	8.403	-0.029	□

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	BC	0.070	0.362	0.427	■		BC	-0.020	10.689	-0.006	□
	BD	-0.286	8.164	-0.101	□		BD	0.462	4.876	0.341	■
	PC	0.200	11.444	0.030	□		PC	-0.333	10.550	-0.089	□
	MT	0.034	0.998	0.123	□		MT	-0.014	0.686	-0.064	□
	XT	0.068	1.659	0.111	□		XT	0.056	0.927	0.246	□
	PT	-0.500	7.221	-0.231	□		PT	0.300	9.304	0.101	□
	AT	-0.643	8.734	-0.332	■		AT	-0.667	5.598	-0.473	■
	ON	0.000	1.393	0.197	□		ON	-0.500	7.445	-0.207	□
	OF	-0.500	9.059	-0.244	□		OF	0.333	12.077	0.124	□
	DR	-0.545	8.514	-0.220	□		DR	0.833	18.414	0.202	□
GSL CB	MC	0.029	0.326	0.392	■	**LE CB**	MC	0.040	0.323	0.649	■
	XC	0.096	0.378	0.614	■		XC	0.166	0.807	0.696	■
	BC	0.307	1.830	0.520	■		BC	0.154	0.599	0.719	■
	BD	0.667	13.106	0.249	□		BD	-0.750	4.270	-0.435	■
	PC	0.167	6.438	0.291	□		PC	0.175	5.349	0.178	□
	MT	0.030	0.836	0.170	□		MT	0.037	0.581	0.240	□
	XT	0.064	0.827	0.333	■		XT	0.017	0.781	0.070	□
	PT	0.250	6.603	0.152	□		PT	-0.400	11.452	-0.153	□
	AT	0.000	4.709	0.042	□		AT	0.000	21.063	0.012	□
	ON	-0.083	5.825	-0.060	□		ON	-1.077	16.236	-0.314	■
	OF	-0.300	8.632	-0.195	□		OF	-0.818	26.510	-0.130	□
	DR	-0.200	9.745	-0.136	□		DR	0.000	21.494	0.000	□
GSL EB	MC	0.099	0.558	0.485	■	**LE EB**	MC	0.022	0.143	0.520	■
	XC	0.227	1.007	0.497	■		XC	0.077	0.439	0.626	■
	BC	0.192	1.929	0.287	■		BC	0.129	0.959	0.626	■
	BD	-0.091	13.141	-0.096	□		BD	-0.500	35.873	-0.202	□
	PC	0.167	13.564	0.029	□		PC	2.500	34.409	0.340	■
	MT	0.020	0.867	0.064	□		MT	0.029	0.599	0.158	□
	XT	0.069	1.395	0.205	□		XT	0.008	0.788	0.094	□
	PT	0.333	6.170	0.175	□		PT	-0.500	12.270	-0.153	□
	AT	-0.533	6.501	-0.304	■		AT	0.000	18.898	0.048	□
	ON	0.000	3.464	-0.024	□		ON	-1.000	23.785	-0.188	□
	OF	-0.222	9.615	-0.059	□		OF	-1.200	35.191	-0.199	□
	DR	0.000	9.375	-0.030	□		DR	-0.333	23.289	-0.096	□
GSL NB	MC	-0.058	1.694	-0.181	□	**LE WB**	MC	0.079	0.890	0.368	■
	XC	-0.056	3.071	-0.076	□		XC	0.216	3.550	0.404	■
	BC	0.313	0.970	0.614	■		BC	0.206	1.078	0.450	■
	BD	-1.100	16.552	-0.288	■		BD	-1.925	11.060	-0.556	■
	PC	0.273	20.109	0.070	□		PC	-0.875	8.139	-0.408	■
	MT	0.030	0.618	0.170	□		MT	0.030	0.581	0.181	□
	XT	0.070	0.624	0.399	■		XT	0.021	0.886	0.094	□
	PT	-0.118	7.870	-0.054	□		PT	0.000	18.224	0.000	□
	AT	-0.083	4.389	-0.084	□		AT	0.556	20.029	0.100	□
	ON	0.286	11.957	0.101	□		ON	0.333	12.809	0.065	□
	OF	-0.250	7.792	-0.174	□		OF	-0.125	23.980	-0.012	□
	DR	-0.875	11.376	-0.282	■		DR	-0.286	23.870	-0.053	□
GSL AB	MC	0.040	0.303	0.532	■	**LE AB**	MC	0.039	0.313	0.462	■
	XC	0.094	0.410	0.556	■		XC	0.151	0.798	0.673	■
	BC	0.301	1.678	0.345	■		BC	0.129	0.577	0.696	■
	BD	0.500	12.994	0.254	□		BD	-1.333	7.061	-0.649	■
	PC	0.200	5.658	0.202	□		PC	0.385	8.225	0.183	□

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	MT	0.049	0.919	0.170	□		MT	0.031	0.577	0.216	□
	XT	0.080	0.863	0.328	■		XT	0.027	0.769	0.124	□
	PT	0.286	5.940	0.148	□		PT	-0.462	13.307	-0.153	□
	AT	0.000	4.784	-0.042	□		AT	0.000	20.116	-0.012	□
	ON	-0.059	5.047	-0.066	□		ON	-0.556	15.015	-0.124	□
	OF	-0.286	7.815	-0.161	□		OF	-0.455	28.865	-0.118	□
	DR	-0.100	9.212	-0.060	□		DR	-0.143	21.125	-0.060	□
	MC	0.078	0.454	0.427	■						
	XC	0.186	0.581	0.649	■						
	BC	0.510	1.843	0.462	■						
	BD	1.167	12.440	0.312	■						
	PC	0.615	9.289	0.235	□						
LA AB	MT	0.032	0.757	0.135	□						
	XT	0.076	1.215	0.275	□						
	PT	0.200	7.632	0.113	□						
	AT	0.000	9.680	-0.030	□						
	ON	-0.167	6.815	-0.072	□						
	OF	-0.400	9.050	-0.119	□						
	DR	-0.500	12.486	-0.152	□						

Table B.7.4. Fall time-series results for each parameter where β = Theil-Sen slope, s_e = standard error of a Theil-Sen regression, τ = Mann-Kendall tau value, and p = Mann-Kendall p-value. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	MC	0.026	0.212	0.357	■		MC	0.238	3.369	0.170	□
	XC	0.063	0.347	0.462	■		XC	0.055	3.737	0.041	□
	BC	0.158	0.884	0.579	■		BC	-0.124	5.104	-0.135	□
	BD	0.040	2.666	0.085	□		BD	0.222	3.658	0.242	□
	PC	0.000	6.427	0.048	□		PC	0.077	6.525	0.047	□
GBL CB	MT	0.037	0.847	0.146	□	LW NB	MT	-0.020	1.260	-0.029	□
	XT	0.043	1.123	0.176	□		XT	0.023	1.801	0.059	□
	PT	0.000	4.439	0.030	□		PT	0.200	2.974	0.292	■
	AT	0.000	4.298	-0.062	□		AT	0.000	3.869	0.030	□
	ON	0.000	1.237	0.132	□		ON	-0.455	7.494	-0.196	□
	OF	-0.500	8.783	-0.236	□		OF	0.143	12.827	0.072	□
	DR	-0.600	8.545	-0.232	□		DR	0.833	17.968	0.219	□
	MC	0.202	1.172	0.481	■		MC	-0.017	1.657	-0.076	□
	XC	0.367	1.583	0.587	■		XC	0.056	2.867	0.099	□
	BC	0.216	1.224	0.493	■		BC	-0.096	6.651	-0.099	□
	BD	0.400	5.012	0.427	■		BD	0.250	3.940	0.266	□
GBL SB	PC	-0.375	5.998	-0.212	□	LW SB	PC	0.727	6.519	0.375	■
	MT	0.066	1.040	0.240	□		MT	0.000	1.453	0.006	□
	XT	0.052	1.327	0.141	□		XT	-0.030	1.584	-0.029	□
	PT	0.133	4.888	0.273	□		PT	0.077	3.587	0.180	□
	AT	0.000	4.102	0.048	□		AT	0.000	4.454	-0.006	□
	ON	0.000	1.572	0.108	□		ON	-0.385	7.704	-0.155	□
	OF	-0.286	7.583	-0.173	□		OF	0.188	12.658	0.064	□

Lake	Parameter	β	s_e	τ	p	Lake	Parameter	β	s_e	τ	p
	DR	-0.462	7.106	-0.218	□		DR	0.400	18.794	0.130	□
GBL AB	MC	0.027	0.233	0.380	■	LW AB	MC	0.092	2.709	0.135	□
	XC	0.069	0.376	0.474	■		XC	0.000	3.257	-0.006	□
	BC	0.173	0.989	0.544	■		BC	-0.146	5.387	-0.111	□
	BD	0.000	3.922	-0.049	□		BD	0.235	3.145	0.309	■
	PC	0.000	6.427	0.048	□		PC	0.200	6.132	0.101	□
	MT	0.033	0.849	0.146	□		MT	-0.012	1.260	-0.006	□
	XT	0.041	1.139	0.176	□		XT	0.023	1.790	0.053	□
	PT	0.000	4.439	0.030	□		PT	0.200	3.093	0.308	■
	AT	0.000	4.678	-0.048	□		AT	0.000	4.189	0.018	□
	ON	0.000	1.393	0.197	□		ON	-0.500	7.445	-0.207	□
	OF	-0.500	9.059	-0.244	□		OF	0.333	12.077	0.124	□
	DR	-0.545	8.514	-0.220	□		DR	0.833	18.414	0.202	□
GSL CB	MC	0.039	0.466	0.298	■	LE CB	MC	0.093	3.000	0.333	■
	XC	0.038	0.422	0.392	■		XC	0.127	5.956	0.287	■
	BC	0.089	0.508	0.520	■		BC	0.054	7.613	0.076	□
	BD	-0.250	2.850	-0.306	■		BD	0.000	4.434	0.037	□
	PC	-0.280	8.655	-0.207	□		PC	0.617	12.146	0.375	■
	MT	0.025	0.950	0.158	□		MT	0.031	1.066	0.181	□
	XT	0.071	1.347	0.146	□		XT	0.080	1.572	0.170	□
	PT	-0.333	3.109	-0.365	■		PT	-0.118	3.483	-0.198	□
	AT	-0.200	3.842	-0.163	□		AT	-0.200	4.519	-0.179	□
	ON	-0.083	5.825	-0.060	□		ON	-1.077	16.236	-0.314	■
	OF	-0.300	8.632	-0.195	□		OF	-0.818	26.510	-0.130	□
	DR	-0.200	9.745	-0.136	□		DR	0.000	21.494	0.000	□
GSL EB	MC	0.096	0.847	0.450	■	LE EB	MC	0.038	3.094	0.357	■
	XC	0.169	1.269	0.509	■		XC	0.065	6.067	0.322	■
	BC	0.113	0.667	0.509	■		BC	0.084	7.712	0.205	□
	BD	0.200	4.363	0.336	■		BD	0.277	6.548	0.136	□
	PC	-0.250	9.659	-0.112	□		PC	0.367	10.288	0.152	□
	MT	0.028	0.890	0.123	□		MT	0.027	0.921	0.135	□
	XT	0.066	1.342	0.135	□		XT	0.083	1.765	0.229	□
	PT	-0.154	3.386	-0.170	□		PT	-0.200	5.426	-0.224	□
	AT	-0.133	2.263	-0.198	□		AT	0.000	4.109	0.018	□
	ON	0.000	3.464	-0.024	□		ON	-1.000	23.785	-0.188	□
	OF	-0.222	9.615	-0.059	□		OF	-1.200	35.191	-0.199	□
	DR	0.000	9.375	-0.030	□		DR	-0.333	23.289	-0.096	□
GSL NB	MC	-0.044	2.562	-0.064	□	LE WB	MC	-0.041	1.336	-0.357	■
	XC	0.138	2.738	0.170	□		XC	-0.038	2.801	-0.146	□
	BC	0.155	2.422	0.158	□		BC	-0.006	8.770	-0.006	□
	BD	-0.214	6.416	-0.117	□		BD	-0.364	6.160	-0.283	■
	PC	0.063	8.540	0.041	□		PC	-0.235	8.501	-0.089	□
	MT	0.061	1.697	0.170	□		MT	0.016	1.029	0.123	□
	XT	0.134	2.524	0.205	□		XT	0.106	1.902	0.240	□
	PT	-0.267	5.490	-0.193	□		PT	-0.167	4.538	-0.259	□
	AT	0.000	1.183	-0.145	□		AT	0.000	1.970	-0.146	□
	ON	0.286	11.957	0.101	□		ON	0.333	12.809	0.065	□
	OF	-0.250	7.792	-0.174	□		OF	-0.125	23.980	-0.012	□
	DR	-0.875	11.376	-0.282	■		DR	-0.286	23.870	-0.053	□
GSL AB	MC	0.047	0.413	0.322	■	LE AB	MC	0.047	2.679	0.228	□
	XC	0.048	0.401	0.544	■		XC	0.083	5.473	0.275	□

Lake	Parameter	β	s$_e$	τ	p	Lake	Parameter	β	s$_e$	τ	p
	BC	0.090	0.582	0.544	■		BC	0.058	7.683	0.053	□
	BD	-0.111	2.567	-0.149	□		BD	0.000	4.557	-0.032	□
	PC	-0.364	8.624	-0.243	□		PC	0.444	11.565	0.169	□
	MT	0.029	0.965	0.135	□		MT	0.023	1.006	0.251	□
	XT	0.071	1.313	0.205	□		XT	0.093	1.676	0.251	□
	PT	-0.250	2.847	-0.291	□		PT	-0.100	3.552	-0.156	□
	AT	-0.200	3.678	-0.151	□		AT	-0.182	4.291	-0.189	□
	ON	-0.059	5.047	-0.066	□		ON	-0.556	15.015	-0.124	□
	OF	-0.286	7.815	-0.161	□		OF	-0.455	28.865	-0.118	□
	DR	-0.100	9.212	-0.060	□		DR	-0.143	21.125	-0.060	□
LA AB	MC	0.108	0.597	0.567	■						
	XC	0.186	0.836	0.602	■						
	BC	0.243	2.139	0.333	■						
	BD	-0.077	5.968	-0.042	□						
	PC	-0.583	7.017	-0.415	■						
	MT	0.043	1.161	0.123	□						
	XT	0.059	1.785	0.099	□						
	PT	-0.400	4.603	-0.317	■						
	AT	-0.176	3.091	-0.245	□						
	ON	-0.167	6.815	-0.072	□						
	OF	-0.400	9.050	-0.119	□						
	DR	-0.500	12.486	-0.152	□						

Table B.7.5. Annual Vector Autoregression results for each parameter where AICc = Akaike information criterion second order, fpe = final prediction error, and s$_e$ = standard error. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	Adj. r2	Median Adj. r^2	AICc	fpe	s$_e$	Min-Max Param
GBL CB	MT	-0.20 - 0.83	0.10	116.73 - 143.71	0.18 - 1.05	0.43 - 0.99	bd - bc
	XT	-0.13 - 0.84	0.19	115.28 - 142.44	0.17 - 0.99	0.41 - 0.95	bd - bc
	AT	-0.01 - 0.80	0.51	118.88 - 141.11	0.20 - 1.17	0.46 - 0.90	bd - bc
	PT	-0.36 - 0.80	0.30	119.77 - 143.55	0.13 - 2.11	0.46 - 0.98	pc - bc
	ON	-0.14 - 0.78	0.04	120.67 - 144.95	0.20 - 1.84	0.48 - 1.03	bd - bc
	OF	-0.12 - 0.79	0.18	119.60 - 142.64	0.19 - 0.39	0.47 - 0.96	pc - bc
	DR	-0.10 - 0.79	0.19	119.93 - 142.70	0.14 - 0.45	0.47 - 0.96	pc - bc
GBL SB	MT	0.04 - 0.96	0.90	94.99 - 139.19	0.04 - 1.20	0.22 - 0.86	pc - mc
	XT	0.07 - 0.95	0.93	97.96 - 138.51	0.04 - 1.35	0.24 - 0.84	pc - mc
	AT	-0.08 - 0.98	0.91	80.34 - 140.49	0.03 - 2.06	0.14 - 0.90	pc - mc
	PT	-0.50 - 0.95	0.94	98.18 - 146.30	0.02 - 2.20	0.24 - 1.07	pc - bc
	ON	-0.26 - 0.95	0.92	95.91 - 141.77	0.07 - 2.38	0.23 - 0.96	pc - mc
	OF	-0.58 - 0.95	0.91	98.29 - 147.32	0.04 - 1.91	0.24 - 1.10	pc - mc
	DR	-0.60 - 0.94	0.91	98.67 - 147.50	0.04 - 1.77	0.24 - 1.11	pc - mc
GBL AB	MT	-0.12 - 0.82	0.14	117.32 - 142.72	0.24 - 1.00	0.43 - 0.99	bd - bc
	XT	-0.07 - 0.81	0.26	118.53 - 140.56	0.19 - 0.75	0.45 - 0.90	bd - bc
	AT	0.30 - 0.82	0.48	117.29 - 134.86	0.12 - 0.55	0.43 - 0.76	bd - bc
	PT	-0.33 - 0.76	0.20	122.66 - 144.24	0.18 - 1.77	0.51 - 1.01	bd - bc

	ON	-0.05 - 0.74	0.23	123.55 - 145.54	0.23 - 0.87	0.52 - 1.04	pc - bc
	OF	-0.05 - 0.75	0.14	122.56 - 143.28	0.22 - 0.83	0.51 - 0.97	pc - bc
	DR	-0.04 - 0.75	0.15	122.48 - 143.13	0.12 - 0.80	0.51 - 0.97	pc - bc
GSL CB	MT	-0.15 - 0.83	0.82	113.82 - 144.61	0.04 - 0.79	0.39 - 1.01	bd - bc
	XT	-0.20 - 0.79	0.64	120.36 - 144.29	0.48 - 1.04	0.47 - 1.00	pc - bc
	AT	0.28 - 0.85	0.71	113.90 - 136.10	0.07 - 0.73	0.39 - 0.77	pc - bc
	PT	-0.17 - 0.83	0.66	115.93 - 143.07	0.44 - 1.86	0.42 - 0.97	pc - bc
	ON	-0.23 - 0.78	0.65	120.82 - 145.19	0.54 - 2.24	0.48 - 1.03	pc - bc
	OF	0.09 - 0.82	0.64	114.19 - 140.00	0.22 - 1.23	0.40 - 0.87	pc - mc
	DR	-0.09 - 0.86	0.65	109.79 - 141.86	0.12 - 2.19	0.35 - 0.93	pc - mc
GSL EB	MT	-0.09 - 0.91	0.82	107.14 - 147.24	0.14 - 2.62	0.31 - 1.11	bd - xc
	XT	-0.28 - 0.92	0.81	106.32 - 149.49	0.07 - 2.48	0.31 - 1.19	bd - xc
	AT	-0.35 - 0.91	0.80	107.01 - 151.14	0.16 - 2.55	0.32 - 1.24	bd - xc
	PT	-0.52 - 0.96	0.89	95.05 - 152.67	0.01 - 0.71	0.21 - 1.31	bd - xc
	ON	-0.60 - 0.92	0.79	106.29 - 154.10	0.11 - 3.33	0.31 - 1.36	bd - xc
	OF	-0.45 - 0.91	0.77	107.87 - 152.28	0.09 - 1.67	0.32 - 1.29	bd - xc
	DR	-0.34 - 0.91	0.76	108.52 - 150.58	0.06 - 1.51	0.33 - 1.23	bd - xc
GSL NB	MT	-0.75 - 0.71	0.47	126.41 - 153.39	0.46 - 3.08	0.57 - 1.33	pc - bc
	XT	-0.80 - 0.74	0.36	124.48 - 153.94	0.71 - 4.08	0.54 - 1.35	pc - bc
	AT	-0.72 - 0.77	0.51	122.61 - 152.65	0.29 - 5.03	0.51 - 1.31	pc - bc
	PT	-0.53 - 0.78	0.34	121.37 - 151.24	0.82 - 6.23	0.49 - 1.25	pc - bc
	ON	-0.05 - 0.80	0.40	120.48 - 145.42	0.45 - 2.99	0.48 - 1.03	pc - bc
	OF	-0.37 - 0.71	0.51	126.46 - 149.44	0.54 - 5.09	0.57 - 1.18	pc - bc
	DR	-0.16 - 0.78	0.46	121.77 - 147.05	0.05 - 1.03	0.49 - 1.09	pc - bc
GSL AB	MT	-0.11 - 0.86	0.79	112.11 - 141.85	0.03 - 1.61	0.37 - 0.93	pc - bc
	XT	-0.59 - 0.84	0.70	116.43 - 148.59	0.22 - 3.53	0.42 - 1.14	pc - bc
	AT	-0.26 - 0.89	0.67	110.20 - 145.96	0.05 - 1.44	0.34 - 1.05	bd - bc
	PT	-0.53 - 0.86	0.70	113.69 - 146.37	0.16 - 3.19	0.38 - 1.07	pc - bc
	ON	-0.48 - 0.83	0.70	116.99 - 148.47	0.33 - 6.30	0.43 - 1.14	bd - bc
	OF	-0.39 - 0.84	0.70	111.87 - 147.57	0.21 - 2.37	0.39 - 1.11	bd - bc
	DR	-0.53 - 0.89	0.70	107.74 - 146.02	0.07 - 2.75	0.32 - 1.05	pc - mc
LA	MT	0.13 - 0.71	0.45	125.61 - 140.70	0.09 - 0.27	0.56 - 0.90	pc - mc
	XT	-0.18 - 0.62	0.47	128.53 - 145.17	0.30 - 1.99	0.61 - 1.05	pc - mc
	AT	-0.59 - 0.48	0.15	134.77 - 150.70	0.88 - 2.58	0.74 - 1.22	pc - mc
	PT	-0.72 - 0.48	0.11	134.73 - 151.80	0.55 - 1.96	0.74 - 1.27	pc - mc
	ON	-0.48 - 0.63	0.40	129.58 - 149.66	0.65 - 1.41	0.63 - 1.18	pc - mc
	OF	-0.71 - 0.56	0.28	130.37 - 151.58	0.22 - 0.62	0.67 - 1.27	pc - mc
	DR	-0.55 - 0.53	0.24	133.11 - 150.32	0.20 - 1.36	0.71 - 1.21	pc - mc
LW NB	MT	-0.58 - -0.01	-0.16	141.85 - 147.25	0.37 - 0.76	0.92 - 1.10	pc - mc
	XT	-0.57 - 0.35	0.22	134.25 - 143.79	0.70 - 2.61	0.73 - 1.00	bd - xc
	AT	-0.29 - 0.23	-0.09	133.27 - 149.00	1.00 - 3.81	0.71 - 1.16	bc - bd
	PT	-0.32 - 0.18	0.08	136.33 - 142.57	0.35 - 1.46	0.80 - 0.96	pc - mc
	ON	-0.20 - 0.07	0.00	133.97 - 146.44	0.70 - 1.35	0.73 - 1.07	bd - mc
	OF	-0.53 - 0.00	-0.18	137.98 - 147.29	0.72 - 2.04	0.83 - 1.10	bd - xc
	DR	-0.38 - 0.03	-0.14	138.06 - 146.83	0.56 - 2.17	0.84 - 1.09	bd - xc
LW SB	MT	-0.24 - 0.50	0.06	132.59 - 147.65	1.41 - 8.10	0.71 - 1.11	bc - mc
	XT	-0.40 - 0.08	-0.17	139.83 - 150.55	0.55 - 1.97	0.88 - 1.22	mc - bc
	AT	-0.53 - 0.20	0.11	136.06 - 152.28	1.67 - 7.09	0.78 - 1.28	xc - bd
	PT	-0.34 - 0.18	-0.12	141.15 - 150.34	1.51 - 3.02	0.91 - 1.21	xc - mc
	ON	-0.47 - 0.24	-0.16	134.85 - 150.91	0.48 - 4.69	0.75 - 1.25	xc - pc
	OF	-0.49 - 0.17	-0.27	136.24 - 152.05	1.81 - 4.77	0.78 - 1.27	xc - pc
	DR	-0.48 - 0.13	-0.27	136.82 - 151.86	1.08 - 4.10	0.80 - 1.26	xc - pc

266

Lake	Parameter	Adj. r2	Median Adj. r²	AICc	fpe	s_e	Min-Max Param
LW AB	MT	-0.42 - 0.19	-0.16	138.32 - 147.20	0.98 - 1.77	0.83 - 1.09	pc - xc
	XT	-0.43 - 0.36	0.08	134.32 - 142.75	0.62 - 2.57	0.74 - 0.96	pc - xc
	AT	-0.17 - 0.30	0.07	133.28 - 144.63	0.50 - 1.52	0.72 - 1.01	pc - xc
	PT	-0.43 - 0.37	0.23	134.86 - 142.13	0.55 - 1.94	0.76 - 0.96	bd - mc
	ON	-0.53 - 0.21	-0.09	133.81 - 146.45	0.91 - 2.10	0.74 - 1.07	bd - xc
	OF	-0.61 - 0.14	-0.21	138.71 - 147.22	0.97 - 2.45	0.84 - 1.11	bd - xc
	DR	-0.56 - 0.22	-0.13	137.89 - 146.60	0.72 - 2.37	0.82 - 1.08	bd - xc
LE CB	MT	0.22 - 0.87	0.75	106.06 - 138.52	0.12 - 0.92	0.33 - 0.86	bd - xc
	XT	0.44 - 0.87	0.74	110.43 - 134.20	0.13 - 0.76	0.35 - 0.73	pc - xc
	AT	0.03 - 0.92	0.86	99.72 - 144.12	0.13 - 1.78	0.26 - 0.99	bd - xc
	PT	0.40 - 0.87	0.84	111.92 - 135.61	0.06 - 1.53	0.37 - 0.77	bd - mc
	ON	-0.04 - 0.88	0.73	109.03 - 145.10	0.10 - 0.74	0.33 - 1.03	bd - xc
	OF	0.16 - 0.86	0.77	111.35 - 140.94	0.16 - 1.08	0.36 - 0.91	bd - xc
	DR	0.10 - 0.87	0.78	109.10 - 142.39	0.08 - 1.28	0.34 - 0.95	bd - xc
LE EB	MT	0.25 - 0.62	0.57	123.84 - 140.44	0.27 - 0.89	0.53 - 0.90	bd - bc
	XT	0.42 - 0.63	0.57	123.63 - 134.54	0.16 - 0.73	0.53 - 0.75	pc - bc
	AT	0.14 - 0.87	0.71	105.98 - 142.55	0.11 - 2.37	0.31 - 0.97	bd - bc
	PT	0.38 - 0.87	0.65	113.19 - 138.35	0.03 - 0.95	0.38 - 0.83	bd - mc
	ON	0.57 - 0.70	0.66	119.11 - 132.55	0.17 - 0.55	0.47 - 0.69	bd - bc
	OF	0.27 - 0.64	0.61	123.19 - 138.86	0.09 - 0.74	0.53 - 0.88	bd - mc
	DR	0.24 - 0.66	0.56	120.10 - 140.82	0.24 - 0.49	0.49 - 0.91	bd - bc
LE WB	MT	0.12 - 0.67	0.16	124.95 - 141.16	0.82 - 2.34	0.55 - 0.91	mc - bc
	XT	0.27 - 0.71	0.56	123.23 - 138.46	0.40 - 1.69	0.52 - 0.83	mc - bc
	AT	0.10 - 0.79	0.30	118.41 - 142.00	0.47 - 2.36	0.44 - 0.93	xc - bc
	PT	0.10 - 0.77	0.30	119.86 - 142.05	0.13 - 1.04	0.47 - 0.93	mc - bc
	ON	0.09 - 0.64	0.19	126.53 - 141.02	0.29 - 2.21	0.58 - 0.90	pc - bc
	OF	0.06 - 0.62	0.48	124.69 - 142.70	0.38 - 1.14	0.55 - 0.95	mc - bc
	DR	0.25 - 0.65	0.33	126.13 - 138.94	0.21 - 0.67	0.57 - 0.85	mc - bc
LE AB	MT	0.43 - 0.84	0.75	112.67 - 136.27	0.26 - 0.74	0.39 - 0.78	bd - xc
	XT	0.54 - 0.84	0.74	113.31 - 132.26	0.15 - 0.43	0.39 - 0.69	pc - xc
	AT	0.44 - 0.92	0.87	101.08 - 136.01	0.17 - 1.07	0.26 - 0.77	bd - xc
	PT	0.64 - 0.89	0.83	109.07 - 128.94	0.02 - 0.67	0.33 - 0.62	pc - mc
	ON	0.41 - 0.87	0.81	110.82 - 136.51	0.25 - 0.95	0.35 - 0.79	bd - xc
	OF	0.40 - 0.85	0.76	112.24 - 136.87	0.16 - 0.62	0.38 - 0.80	bd - xc
	DR	0.37 - 0.86	0.79	111.55 - 137.70	0.09 - 0.37	0.36 - 0.82	bd - xc

Table B.7.6. Spring Autoregression results for each parameter where AICc = Akaike information criterion second order, fpe = final prediction error, and s_e = standard error. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	Adj. r2	Median Adj. r²	AICc	fpe	s_e	Min-Max Param
GBL CB	MT	0.26 - 0.91	0.81	105.57 - 142.00	0.13 - 0.94	0.30 - 0.93	bd - bc
	XT	0.31 - 0.95	0.87	90.34 - 140.81	0.07 - 1.22	0.20 - 0.89	bd - bc
	AT	0.18 - 0.91	0.82	107.64 - 143.20	0.08 - 0.97	0.32 - 0.97	bd - mc
	PT	0.38 - 0.92	0.85	104.97 - 138.94	0.23 - 1.02	0.29 - 0.84	bd - mc
	ON	0.29 - 0.91	0.80	105.66 - 137.15	0.16 - 1.50	0.30 - 0.80	pc - bc
	OF	0.29 - 0.90	0.81	105.41 - 140.97	0.07 - 0.52	0.33 - 0.90	bd - mc

	DR	0.28 - 0.91	0.81	105.39 - 140.89	0.07 - 0.52	0.32 - 0.90	bd - mc
GBL SB	MT	-0.02 - 0.69	0.03	128.53 - 146.32	0.07 - 0.61	0.61 - 1.07	pc - bc
	XT	0.01 - 0.65	0.18	130.42 - 146.06	0.20 - 0.74	0.65 - 1.07	bd - bc
	AT	0.03 - 0.66	0.31	129.22 - 146.39	0.10 - 1.49	0.63 - 1.07	bd - bc
	PT	0.04 - 0.66	0.29	129.77 - 144.64	0.28 - 0.65	0.63 - 1.01	xc - bc
	ON	-0.33 - 0.65	-0.05	129.60 - 151.54	0.73 - 2.88	0.63 - 1.25	bd - bc
	OF	-0.05 - 0.59	0.12	132.92 - 146.77	0.49 - 1.17	0.70 - 1.10	bd - bc
	DR	-0.04 - 0.62	0.08	131.26 - 147.28	0.42 - 0.89	0.67 - 1.10	bd - bc
GBL AB	MT	0.31 - 0.87	0.84	114.59 - 139.25	0.11 - 0.93	0.39 - 0.86	pc - mc
	XT	0.40 - 0.89	0.88	111.25 - 137.14	0.07 - 0.78	0.36 - 0.80	pc - bc
	AT	0.56 - 0.91	0.86	107.35 - 131.13	0.10 - 0.65	0.32 - 0.67	pc - mc
	PT	0.47 - 0.90	0.87	110.21 - 135.06	0.17 - 1.36	0.35 - 0.75	pc - mc
	ON	0.31 - 0.88	0.85	112.32 - 139.19	0.22 - 2.50	0.37 - 0.85	pc - mc
	OF	0.38 - 0.88	0.84	112.40 - 135.97	0.08 - 0.93	0.37 - 0.80	pc - mc
	DR	0.34 - 0.88	0.84	111.59 - 138.09	0.07 - 0.95	0.37 - 0.83	pc - mc
GSL CB	MT	-0.45 - 0.65	0.31	126.46 - 142.13	0.50 - 1.13	0.60 - 0.94	pc - bd
	XT	-0.34 - 0.62	0.28	126.83 - 143.31	0.58 - 1.66	0.61 - 0.97	pc - bd
	AT	-0.58 - 0.77	0.30	121.28 - 146.53	0.06 - 1.99	0.49 - 1.07	pc - xc
	PT	-0.53 - 0.69	0.43	121.66 - 143.70	0.07 - 0.98	0.53 - 0.98	pc - mc
	ON	-0.50 - 0.60	0.36	129.54 - 143.18	0.38 - 2.00	0.64 - 0.97	pc - bd
	OF	-0.06 - 0.71	0.39	123.25 - 138.23	0.23 - 1.31	0.54 - 0.83	pc - bd
	DR	-0.35 - 0.67	0.42	126.05 - 141.64	0.38 - 1.76	0.58 - 0.92	pc - bd
GSL EB	MT	-0.10 - 0.46	0.31	136.49 - 147.29	0.35 - 2.86	0.78 - 1.10	bc - xc
	XT	0.12 - 0.50	0.39	135.00 - 141.69	0.57 - 1.45	0.75 - 0.93	pc - bd
	AT	0.00 - 0.62	0.33	130.60 - 145.35	0.54 - 1.77	0.65 - 1.03	pc - bd
	PT	0.02 - 0.49	0.25	135.62 - 144.98	0.53 - 2.24	0.76 - 1.03	bc - xc
	ON	-0.33 - 0.53	0.15	133.84 - 150.54	0.38 - 3.89	0.73 - 1.21	bc - xc
	OF	-0.05 - 0.55	0.28	133.29 - 145.63	0.40 - 1.68	0.71 - 1.06	bc - xc
	DR	-0.01 - 0.60	0.29	131.42 - 144.93	0.22 - 1.62	0.67 - 1.04	pc - xc
GSL NB	MT	-0.75 - 0.65	-0.17	124.67 - 153.98	0.62 - 3.46	0.55 - 1.35	xc - bc
	XT	-0.96 - 0.69	-0.10	122.96 - 154.22	0.48 - 5.05	0.51 - 1.36	bd - bc
	AT	-0.50 - 0.45	-0.20	131.44 - 149.88	1.06 - 2.94	0.68 - 1.19	bd - bc
	PT	-0.72 - 0.46	-0.16	131.67 - 152.19	0.32 - 3.65	0.68 - 1.28	bd - bc
	ON	-0.64 - 0.54	-0.08	129.25 - 151.10	1.33 - 4.02	0.62 - 1.24	bd - bc
	OF	-0.78 - 0.45	-0.13	131.67 - 152.51	1.33 - 2.92	0.68 - 1.29	bd - bc
	DR	-0.67 - 0.49	-0.16	130.71 - 151.45	0.60 - 1.80	0.66 - 1.25	bd - bc
GSL AB	MT	-0.33 - 0.74	0.26	122.97 - 140.51	0.15 - 1.21	0.52 - 0.90	pc - bd
	XT	-0.29 - 0.74	0.23	122.72 - 143.81	0.18 - 1.86	0.52 - 0.99	pc - bd
	AT	-0.48 - 0.85	0.30	114.66 - 143.19	0.20 - 2.27	0.40 - 0.96	pc - bd
	PT	-0.53 - 0.77	0.65	119.95 - 143.64	0.08 - 0.43	0.49 - 0.98	pc - xc
	ON	-0.54 - 0.70	0.34	125.27 - 145.33	0.15 - 3.27	0.56 - 1.04	pc - bd
	OF	-0.10 - 0.78	0.36	120.54 - 138.44	0.08 - 1.51	0.48 - 0.83	pc - bd
	DR	-0.31 - 0.78	0.43	120.64 - 144.20	0.17 - 1.67	0.48 - 1.00	pc - bd
LA	MT	-0.79 - 0.83	-0.29	110.05 - 155.32	0.00 - 0.08	0.34 - 1.42	bd - pc
	XT	-0.82 - 0.71	-0.23	116.78 - 155.25	0.03 - 0.46	0.43 - 1.42	bd - pc
	AT	0.08 - 0.38	0.25	127.89 - 142.72	0.45 - 0.70	0.64 - 0.95	mc - bc
	PT	-0.51 - 0.61	0.20	122.30 - 150.46	0.06 - 0.69	0.51 - 1.26	bd - pc
	ON	-0.68 - 0.49	-0.17	123.79 - 154.58	0.53 - 1.69	0.56 - 1.38	bd - pc
	OF	-0.45 - 0.66	0.06	118.65 - 149.84	0.11 - 0.72	0.46 - 1.24	bd - pc
	DR	-0.59 - 0.55	-0.05	122.45 - 152.96	0.30 - 0.82	0.53 - 1.33	bd - pc
LW NB	MT	-0.39 - 0.59	0.41	105.50 - 134.66	0.10 - 0.74	0.30 - 0.74	mc - bc
	XT	-0.51 - 0.55	0.18	106.99 - 134.58	0.11 - 0.53	0.31 - 0.74	mc - bc

268

	AT	-0.15 - 0.52	0.22	101.97 - 131.18	0.33 - 1.08	0.27 - 0.67	mc - bd
	PT	-0.12 - 0.74	0.25	96.31 - 129.90	0.01 - 0.17	0.23 - 0.64	pc - bc
	ON	-0.44 - 0.75	0.11	105.32 - 131.46	0.09 - 0.69	0.30 - 0.67	mc - bc
	OF	-0.33 - 0.57	0.17	103.71 - 132.77	0.10 - 0.36	0.28 - 0.70	xc - bc
	DR	-0.41 - 0.66	0.09	105.46 - 133.26	0.13 - 0.60	0.30 - 0.71	xc - bc
LW SB	MT	0.25 - 0.64	0.51	117.63 - 139.63	0.68 - 2.69	0.44 - 0.90	pc - bc
	XT	-0.35 - 0.62	0.06	120.25 - 143.87	0.38 - 2.26	0.52 - 0.99	mc - bd
	AT	-0.37 - 0.03	-0.11	134.65 - 145.62	1.18 - 3.42	0.76 - 1.06	mc - bd
	PT	-0.29 - 0.38	0.16	131.15 - 143.32	0.63 - 3.30	0.67 - 0.97	mc - pc
	ON	-0.09 - 0.08	-0.02	133.09 - 145.79	1.15 - 4.07	0.72 - 1.08	xc - bd
	OF	-0.30 - 0.14	-0.02	132.40 - 144.47	1.01 - 4.51	0.71 - 1.05	mc - bd
	DR	-0.32 - 0.04	-0.09	133.48 - 146.12	0.77 - 2.82	0.74 - 1.10	mc - bd
LW AB	MT	-0.31 - 0.62	0.28	75.24 - 127.34	0.01 - 0.69	0.12 - 0.60	xc - bc
	XT	-0.25 - 0.45	0.20	73.89 - 129.65	0.02 - 0.82	0.11 - 0.65	xc - bd
	AT	-0.33 - 0.53	0.23	73.53 - 128.08	0.04 - 1.28	0.11 - 0.61	xc - bd
	PT	0.39 - 0.74	0.57	61.61 - 118.35	0.02 - 1.10	0.08 - 0.46	pc - bd
	ON	-0.32 - 0.74	0.27	69.81 - 126.87	0.02 - 0.77	0.10 - 0.58	xc - bd
	OF	-0.43 - 0.55	0.11	75.38 - 130.41	0.03 - 0.93	0.12 - 0.65	xc - bc
	DR	-0.42 - 0.62	0.11	74.86 - 125.31	0.03 - 0.59	0.12 - 0.55	xc - bd
LE CB	MT	-0.06 - 0.35	0.12	132.58 - 143.81	1.00 - 3.49	0.70 - 1.00	xc - bd
	XT	-0.58 - 0.32	-0.38	139.59 - 150.96	1.83 - 4.96	0.87 - 1.23	xc - bd
	AT	-0.89 - 0.20	-0.45	140.82 - 153.86	1.88 - 3.04	0.92 - 1.35	xc - bd
	PT	-0.73 - 0.05	-0.69	145.29 - 152.74	3.51 - 5.18	1.03 - 1.30	xc - bd
	ON	-0.60 - 0.29	-0.40	136.19 - 150.49	1.48 - 5.07	0.78 - 1.21	bc - bd
	OF	-0.57 - 0.24	-0.34	140.85 - 150.70	1.16 - 2.28	0.91 - 1.22	xc - bd
	DR	-0.55 - 0.16	-0.42	143.29 - 149.19	0.47 - 3.10	0.97 - 1.20	xc - bd
LE EB	MT	-0.36 - 0.50	0.09	124.77 - 146.57	0.30 - 1.05	0.54 - 1.07	bc - pc
	XT	-0.59 - 0.37	0.03	129.09 - 148.79	0.56 - 1.57	0.62 - 1.16	bc - mc
	AT	-0.13 - 0.23	0.18	132.47 - 143.24	0.36 - 1.14	0.69 - 0.97	bc - xc
	PT	-0.50 - 0.57	0.21	125.25 - 148.19	0.15 - 1.40	0.56 - 1.13	bc - xc
	ON	-0.76 - 0.43	-0.05	125.74 - 150.98	0.34 - 1.00	0.57 - 1.23	bc - pc
	OF	-0.76 - 0.37	0.15	128.28 - 150.62	0.11 - 0.52	0.61 - 1.22	bc - pc
	DR	-0.68 - 0.43	-0.12	126.02 - 149.03	0.25 - 1.36	0.57 - 1.18	bc - pc
LE WB	MT	-0.38 - 0.02	-0.12	133.47 - 147.63	0.49 - 2.80	0.71 - 1.11	bc - xc
	XT	-0.38 - 0.19	-0.18	134.94 - 147.52	0.46 - 1.67	0.75 - 1.11	bc - xc
	AT	-0.55 - 0.14	-0.11	137.12 - 149.81	0.64 - 3.40	0.80 - 1.19	bc - xc
	PT	-0.30 - 0.31	0.12	128.88 - 147.42	0.70 - 1.82	0.62 - 1.11	bc - bd
	ON	-0.16 - 0.39	0.04	133.42 - 145.14	0.22 - 2.89	0.72 - 1.03	bc - xc
	OF	-0.60 - 0.56	0.04	128.83 - 150.32	0.33 - 2.95	0.63 - 1.21	bc - bd
	DR	-0.45 - 0.47	-0.05	132.54 - 148.77	0.29 - 1.62	0.70 - 1.15	bc - bd
LE AB	MT	0.03 - 0.60	0.19	128.76 - 140.16	0.87 - 3.15	0.63 - 0.88	xc - bd
	XT	-0.53 - 0.37	-0.23	138.84 - 147.22	1.99 - 4.43	0.84 - 1.09	xc - bd
	AT	-0.43 - 0.26	-0.05	138.92 - 144.78	1.53 - 2.20	0.89 - 1.03	xc - bd
	PT	-0.32 - 0.63	-0.23	128.86 - 144.69	0.69 - 3.33	0.63 - 1.02	pc - bd
	ON	-0.79 - 0.40	-0.44	134.79 - 149.50	2.02 - 5.42	0.74 - 1.18	xc - bd
	OF	-0.73 - 0.35	-0.26	137.43 - 148.79	0.81 - 1.68	0.83 - 1.15	xc - bd
	DR	-0.60 - 0.31	-0.36	139.42 - 147.23	0.20 - 2.01	0.87 - 1.11	xc - bd

Table B.7.7. Summer Autoregression results for each parameter where AICc = Akaike information criterion second order, fpe = final prediction error, and s_e = standard error. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD =

Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	Adj. r2	Median Adj. r²	AICc	fpe	s_e	Min-Max Param
GBL CB	MT	0.34 - 0.89	0.48	109.55 - 139.02	0.05 - 0.32	0.34 - 0.86	mc - bc
	XT	0.41 - 0.91	0.51	102.18 - 137.68	0.05 - 0.30	0.30 - 0.82	mc - bc
	AT	0.74 - 0.89	0.77	108.15 - 125.03	0.11 - 0.17	0.33 - 0.55	mc - bd
	PT	0.27 - 0.92	0.53	104.11 - 138.67	0.05 - 0.37	0.29 - 0.87	bd - bc
	ON	0.47 - 0.86	0.65	114.47 - 136.28	0.18 - 0.51	0.39 - 0.78	mc - bc
	OF	0.29 - 0.89	0.55	105.53 - 137.65	0.07 - 0.21	0.32 - 0.85	bd - bc
	DR	0.30 - 0.89	0.54	107.53 - 137.99	0.08 - 0.18	0.33 - 0.85	bd - bc
GBL SB	MT	-0.26 - 0.88	0.86	108.88 - 147.26	0.15 - 1.92	0.34 - 1.10	bd - xc
	XT	-0.16 - 0.90	0.88	107.21 - 145.42	0.14 - 1.65	0.33 - 1.05	bd - bc
	AT	-0.41 - 0.84	0.80	113.13 - 149.28	0.14 - 0.89	0.39 - 1.17	bd - xc
	PT	-0.55 - 0.78	0.75	120.59 - 150.33	0.28 - 2.73	0.48 - 1.22	bd - bc
	ON	0.04 - 0.81	0.78	116.81 - 142.59	0.23 - 2.12	0.43 - 0.95	bd - xc
	OF	-0.23 - 0.82	0.79	116.27 - 146.76	0.19 - 1.30	0.43 - 1.08	bd - bc
	DR	-0.43 - 0.80	0.77	115.54 - 149.48	0.20 - 1.59	0.43 - 1.17	bd - xc
GBL AB	MT	0.33 - 0.89	0.61	110.14 - 139.36	0.06 - 0.24	0.35 - 0.87	mc - bc
	XT	0.42 - 0.90	0.61	100.82 - 137.39	0.05 - 0.17	0.30 - 0.81	mc - bc
	AT	0.74 - 0.84	0.79	114.59 - 125.14	0.09 - 0.19	0.40 - 0.55	mc - bc
	PT	0.40 - 0.90	0.55	107.82 - 137.02	0.02 - 0.24	0.32 - 0.81	bd - bc
	ON	0.43 - 0.82	0.61	117.73 - 137.46	0.15 - 0.39	0.44 - 0.81	mc - bc
	OF	0.43 - 0.87	0.55	108.52 - 137.38	0.04 - 0.18	0.36 - 0.81	mc - bc
	DR	0.42 - 0.86	0.52	109.38 - 137.39	0.04 - 0.15	0.37 - 0.81	mc - bc
GSL CB	MT	0.02 - 0.61	0.48	129.30 - 146.21	0.81 - 2.85	0.62 - 1.06	pc - xc
	XT	-0.02 - 0.58	0.43	130.33 - 146.07	0.80 - 2.73	0.65 - 1.06	bd - xc
	AT	0.02 - 0.61	0.49	129.53 - 146.27	0.97 - 4.25	0.63 - 1.07	pc - xc
	PT	-0.07 - 0.62	0.47	128.71 - 147.80	0.71 - 3.44	0.61 - 1.11	pc - xc
	ON	-0.16 - 0.60	0.36	129.91 - 147.17	0.70 - 4.36	0.64 - 1.10	bd - xc
	OF	0.09 - 0.61	0.47	129.44 - 143.58	0.35 - 1.33	0.63 - 0.98	bd - xc
	DR	0.15 - 0.61	0.47	129.50 - 142.85	0.66 - 2.26	0.63 - 0.97	bd - xc
GSL EB	MT	0.22 - 0.87	0.55	110.88 - 141.21	0.11 - 1.62	0.36 - 0.91	bc - xc
	XT	0.44 - 0.92	0.58	102.46 - 134.91	0.04 - 0.76	0.28 - 0.76	bc - xc
	AT	0.20 - 0.94	0.54	100.35 - 142.56	0.02 - 0.72	0.26 - 0.95	bd - xc
	PT	0.21 - 0.96	0.40	93.47 - 142.45	0.03 - 1.02	0.20 - 0.94	bd - xc
	ON	0.05 - 0.87	0.50	111.36 - 144.65	0.09 - 2.62	0.36 - 1.01	bc - xc
	OF	0.28 - 0.89	0.55	107.02 - 140.13	0.07 - 0.95	0.32 - 0.89	bd - xc
	DR	0.19 - 0.89	0.51	107.70 - 142.88	0.07 - 1.37	0.33 - 0.95	bc - xc
GSL NB	MT	-0.08 - 0.96	0.39	92.99 - 142.91	0.12 - 1.67	0.20 - 0.96	bd - bc
	XT	-0.16 - 0.95	0.27	96.05 - 140.99	0.19 - 2.31	0.22 - 0.91	bd - bc
	AT	-0.32 - 0.95	0.56	96.95 - 142.60	0.23 - 2.12	0.23 - 0.95	bd - bc
	PT	-0.16 - 0.94	0.18	100.67 - 146.96	0.38 - 3.81	0.26 - 1.10	bd - bc
	ON	-0.19 - 0.94	-0.03	100.02 - 147.56	0.39 - 5.06	0.25 - 1.11	bd - bc
	OF	0.06 - 0.94	0.33	99.53 - 143.04	0.29 - 2.99	0.25 - 1.00	pc - bc
	DR	-0.08 - 0.95	0.09	97.57 - 145.69	0.10 - 2.09	0.23 - 1.06	bd - bc
GSL AB	MT	0.09 - 0.65	0.40	127.85 - 142.59	0.61 - 1.96	0.60 - 0.98	pc - xc
	XT	0.02 - 0.64	0.36	128.51 - 144.66	0.60 - 2.07	0.61 - 1.01	bd - xc
	AT	-0.04 - 0.62	0.43	129.21 - 146.69	0.89 - 2.04	0.62 - 1.08	pc - xc
	PT	-0.12 - 0.76	0.48	122.32 - 147.01	0.31 - 3.38	0.50 - 1.09	bd - xc
	ON	-0.14 - 0.64	0.39	128.25 - 147.21	0.93 - 4.77	0.61 - 1.10	bd - xc

	OF	0.09 - 0.65	0.56	127.87 - 143.68	0.34 - 0.98	0.60 - 0.98	bd - xc
	DR	0.18 - 0.64	0.53	128.25 - 142.06	0.56 - 2.19	0.61 - 0.93	bd - xc
LA	MT	0.15 - 0.89	0.41	107.52 - 140.18	0.16 - 1.68	0.32 - 0.89	pc - bc
	XT	0.09 - 0.82	0.39	115.68 - 140.96	0.05 - 1.10	0.42 - 0.91	pc - xc
	AT	0.52 - 0.84	0.64	113.17 - 133.47	0.20 - 1.90	0.39 - 0.72	mc - xc
	PT	0.10 - 0.79	0.39	119.43 - 142.76	0.16 - 1.22	0.46 - 0.96	bd - xc
	ON	0.10 - 0.86	0.54	113.08 - 140.82	0.29 - 1.65	0.38 - 0.91	pc - xc
	OF	0.01 - 0.86	0.49	112.88 - 142.67	0.06 - 0.53	0.38 - 0.96	pc - xc
	DR	0.07 - 0.84	0.46	116.34 - 141.75	0.12 - 1.84	0.42 - 0.93	pc - xc
LW NB	MT	-0.08 - 0.18	0.12	133.30 - 140.86	0.52 - 0.97	0.74 - 0.90	bd - xc
	XT	-0.20 - 0.42	0.21	130.54 - 140.05	0.19 - 1.97	0.68 - 0.88	pc - bc
	AT	-0.44 - 0.47	0.25	132.22 - 141.07	0.25 - 1.00	0.69 - 0.90	bd - xc
	PT	-0.40 - 0.69	0.30	117.93 - 142.64	0.32 - 0.74	0.50 - 0.95	pc - mc
	ON	-0.29 - 0.35	0.15	131.54 - 140.84	0.50 - 1.10	0.67 - 0.90	pc - mc
	OF	-0.49 - 0.40	0.24	134.19 - 143.81	0.28 - 1.56	0.73 - 0.98	pc - xc
	DR	-0.35 - 0.34	0.21	131.47 - 141.78	0.13 - 1.19	0.69 - 0.93	pc - xc
LW SB	MT	-0.50 - 0.64	-0.03	128.18 - 153.09	0.38 - 1.45	0.61 - 1.31	bd - pc
	XT	-0.28 - 0.40	0.01	135.39 - 150.07	0.40 - 1.89	0.77 - 1.20	xc - pc
	AT	-0.13 - 0.49	0.20	133.73 - 148.25	0.55 - 9.89	0.72 - 1.13	bd - pc
	PT	-0.50 - 0.49	0.16	131.03 - 153.00	1.22 - 9.81	0.70 - 1.31	bd - pc
	ON	-0.48 - 0.53	-0.07	129.97 - 152.75	0.95 - 6.54	0.67 - 1.30	bd - pc
	OF	-0.36 - 0.37	-0.09	137.14 - 151.27	0.52 - 4.26	0.80 - 1.24	xc - pc
	DR	-0.45 - 0.43	-0.01	135.56 - 152.49	0.33 - 3.73	0.76 - 1.29	bd - pc
LW AB	MT	0.05 - 0.30	0.22	132.71 - 138.30	0.38 - 0.93	0.70 - 0.83	bd - mc
	XT	-0.02 - 0.45	0.33	129.80 - 139.05	0.16 - 1.72	0.66 - 0.85	pc - xc
	AT	-0.06 - 0.65	0.58	124.45 - 139.93	0.32 - 0.56	0.54 - 0.87	pc - xc
	PT	-0.32 - 0.74	0.52	116.91 - 143.41	0.33 - 1.73	0.45 - 0.97	pc - mc
	ON	-0.18 - 0.40	0.19	131.82 - 141.86	0.50 - 1.56	0.69 - 0.92	bd - mc
	OF	-0.26 - 0.36	0.24	133.74 - 142.67	0.29 - 1.61	0.73 - 0.95	pc - mc
	DR	-0.18 - 0.33	0.25	134.61 - 141.42	0.17 - 1.24	0.75 - 0.92	pc - mc
LE CB	MT	0.00 - 0.72	0.49	122.63 - 146.16	0.38 - 0.90	0.52 - 1.07	pc - bc
	XT	-0.46 - 0.80	0.67	117.65 - 152.36	0.30 - 1.54	0.44 - 1.29	pc - bc
	AT	-0.19 - 0.75	0.56	121.06 - 147.43	0.35 - 6.54	0.49 - 1.14	pc - bc
	PT	-0.46 - 0.78	0.64	118.99 - 152.01	0.22 - 0.46	0.46 - 1.28	pc - bc
	ON	0.13 - 0.74	0.56	121.70 - 143.50	0.33 - 1.62	0.50 - 0.99	pc - bc
	OF	-0.29 - 0.87	0.59	111.48 - 150.52	0.15 - 3.61	0.36 - 1.21	pc - bc
	DR	-0.58 - 0.88	0.61	110.10 - 153.64	0.21 - 3.70	0.34 - 1.34	pc - bc
LE EB	MT	-0.02 - 0.67	0.54	125.88 - 145.43	0.34 - 1.20	0.57 - 1.05	pc - xc
	XT	-0.49 - 0.69	0.46	125.04 - 151.90	0.22 - 2.22	0.56 - 1.27	pc - bc
	AT	-0.48 - 0.69	0.32	123.05 - 152.01	0.27 - 2.20	0.55 - 1.27	pc - bc
	PT	-0.43 - 0.75	0.59	120.47 - 151.52	0.16 - 0.98	0.50 - 1.25	pc - xc
	ON	-0.31 - 0.73	0.54	122.16 - 149.83	0.22 - 1.28	0.51 - 1.19	pc - bc
	OF	-0.16 - 0.87	0.52	110.67 - 147.91	0.01 - 1.10	0.36 - 1.12	pc - bc
	DR	-0.12 - 0.81	0.57	115.66 - 146.81	0.20 - 1.36	0.42 - 1.09	pc - bc
LE WB	MT	-0.20 - 0.68	0.54	123.20 - 143.67	0.27 - 2.18	0.52 - 0.99	pc - bc
	XT	0.10 - 0.66	0.52	126.17 - 139.61	0.21 - 0.90	0.58 - 0.86	pc - bd
	AT	0.30 - 0.64	0.46	125.56 - 137.35	0.73 - 1.22	0.56 - 0.80	pc - bc
	PT	-0.13 - 0.56	0.45	130.41 - 142.41	0.26 - 1.13	0.65 - 0.96	pc - bd
	ON	-0.25 - 0.61	0.47	127.16 - 144.24	0.29 - 1.52	0.58 - 1.01	pc - bc
	OF	-0.31 - 0.61	0.43	127.05 - 145.42	0.31 - 2.18	0.58 - 1.04	pc - bc
	DR	-0.24 - 0.58	0.48	128.34 - 144.51	0.13 - 1.84	0.61 - 1.01	pc - bc
LE A	MT	-0.18 - 0.72	0.63	122.87 - 146.00	0.35 - 1.32	0.52 - 1.06	pc - bd

	XT	0.09 - 0.80	0.64	116.24 - 142.08	0.23 - 0.78	0.43 - 0.93	pc - bc
	AT	-0.27 - 0.77	0.67	118.71 - 147.42	0.50 - 2.86	0.49 - 1.10	pc - bd
	PT	-0.15 - 0.76	0.64	118.85 - 145.62	0.29 - 0.67	0.47 - 1.04	pc - bc
	ON	-0.37 - 0.72	0.60	125.87 - 148.57	0.71 - 1.77	0.56 - 1.14	pc - bd
	OF	-0.22 - 0.88	0.79	109.55 - 146.59	0.13 - 1.75	0.36 - 1.07	pc - bd
	DR	-0.28 - 0.84	0.81	113.05 - 147.36	0.17 - 0.87	0.39 - 1.10	pc - bc

Table B.7.8. Fall Autoregression results for each parameter where AICc = Akaike information criterion second order, fpe = final prediction error, and s_e = standard error. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.

Lake	Parameter	Adj. r2	Median Adj. r²	AICc	fpe	s_e	Min-Max Param
GBL CB	MT	-0.04 - 0.77	0.71	121.26 - 143.91	0.35 - 1.10	0.50 - 0.99	pc - xc
	XT	0.07 - 0.77	0.74	120.88 - 143.36	0.25 - 0.84	0.49 - 0.97	bd - xc
	AT	-0.27 - 0.76	0.63	121.47 - 147.03	0.17 - 0.45	0.49 - 1.10	pc - bc
	PT	-0.02 - 0.86	0.71	112.07 - 143.78	0.12 - 0.79	0.37 - 0.99	pc - bc
	ON	-0.46 - 0.78	0.63	120.05 - 149.67	0.15 - 0.98	0.47 - 1.18	pc - bc
	OF	-0.37 - 0.77	0.67	122.17 - 148.78	0.12 - 0.65	0.50 - 1.15	pc - xc
	DR	-0.40 - 0.77	0.67	122.18 - 148.95	0.09 - 0.79	0.50 - 1.16	pc - xc
GBL SB	MT	-0.42 - 0.73	0.64	120.92 - 149.13	0.58 - 1.69	0.51 - 1.17	pc - xc
	XT	-0.25 - 0.75	0.69	117.35 - 146.93	0.36 - 1.85	0.47 - 1.09	pc - xc
	AT	0.01 - 0.69	0.55	125.20 - 143.65	0.62 - 2.48	0.56 - 0.98	pc - xc
	PT	0.07 - 0.90	0.80	105.61 - 142.97	0.13 - 1.00	0.31 - 0.96	pc - xc
	ON	-0.36 - 0.77	0.65	116.94 - 148.80	0.20 - 1.22	0.46 - 1.15	pc - xc
	OF	-0.29 - 0.69	0.57	125.13 - 147.82	0.30 - 1.20	0.55 - 1.12	pc - xc
	DR	-0.27 - 0.68	0.56	125.93 - 147.58	0.48 - 1.50	0.57 - 1.11	pc - xc
GBL AB	MT	-0.03 - 0.76	0.55	122.91 - 143.74	0.36 - 1.40	0.51 - 0.99	pc - xc
	XT	0.05 - 0.78	0.63	121.52 - 143.00	0.21 - 1.00	0.49 - 0.98	bd - xc
	AT	-0.08 - 0.73	0.60	123.82 - 144.40	0.18 - 0.65	0.53 - 1.01	pc - xc
	PT	-0.01 - 0.79	0.59	119.90 - 143.58	0.11 - 0.77	0.47 - 0.98	pc - xc
	ON	-0.24 - 0.71	0.56	124.14 - 147.16	0.20 - 0.89	0.54 - 1.09	pc - xc
	OF	-0.36 - 0.74	0.42	124.11 - 148.73	0.14 - 0.85	0.53 - 1.15	pc - xc
	DR	-0.34 - 0.74	0.44	123.97 - 148.44	0.12 - 0.70	0.53 - 1.14	pc - xc
GSL CB	MT	-0.53 - 0.88	0.27	111.27 - 152.07	0.26 - 3.27	0.36 - 1.28	mc - bc
	XT	-0.56 - 0.82	0.12	118.12 - 152.08	0.42 - 2.86	0.44 - 1.28	mc - bc
	AT	-0.51 - 0.80	-0.09	119.01 - 151.69	0.50 - 2.74	0.46 - 1.27	mc - bc
	PT	0.08 - 0.82	0.42	117.46 - 143.33	0.36 - 1.47	0.43 - 0.98	mc - bc
	ON	-0.26 - 0.79	0.00	119.66 - 148.43	0.70 - 2.96	0.46 - 1.15	mc - bc
	OF	-0.31 - 0.78	0.05	120.46 - 148.40	0.58 - 2.82	0.48 - 1.16	mc - bc
	DR	-0.24 - 0.79	0.23	119.12 - 144.34	0.38 - 1.63	0.46 - 1.01	bd - bc
GSL EB	MT	0.05 - 0.80	0.38	118.98 - 144.31	0.71 - 1.91	0.46 - 1.00	pc - bc
	XT	0.05 - 0.88	0.46	109.44 - 144.28	0.37 - 1.98	0.35 - 1.00	pc - bc
	AT	-0.01 - 0.88	0.66	111.03 - 143.96	0.25 - 1.24	0.36 - 1.05	bd - bc
	PT	-0.01 - 0.74	0.33	123.49 - 146.97	0.39 - 1.24	0.53 - 1.09	bd - bc
	ON	0.18 - 0.85	0.29	113.71 - 143.15	0.37 - 1.65	0.39 - 0.97	bd - bc
	OF	0.00 - 0.77	0.31	120.29 - 145.20	0.41 - 1.18	0.49 - 1.06	bd - bc
	DR	0.10 - 0.81	0.40	116.72 - 142.54	0.37 - 0.84	0.44 - 0.99	bd - bc
GSL N	MT	-0.06 - 0.49	0.38	131.11 - 140.35	0.77 - 3.30	0.68 - 0.89	mc - bc

	XT	-0.09 - 0.65	0.45	128.09 - 139.98	0.54 - 2.52	0.60 - 0.89	mc - pc
	AT	0.13 - 0.64	0.38	126.71 - 143.49	0.53 - 3.19	0.58 - 0.97	bd - bc
	PT	-0.08 - 0.68	0.52	123.65 - 146.77	0.75 - 3.67	0.54 - 1.08	bd - bc
	ON	-0.06 - 0.71	0.23	121.85 - 145.55	1.07 - 4.78	0.51 - 1.06	mc - bc
	OF	-0.02 - 0.62	0.20	125.77 - 144.40	0.60 - 3.93	0.58 - 1.03	bd - bc
	DR	0.16 - 0.82	0.26	115.10 - 138.67	0.16 - 1.27	0.40 - 0.90	bd - bc
GSL AB	**MT**	-0.20 - 0.87	0.28	112.93 - 148.13	0.39 - 2.69	0.38 - 1.13	mc - bc
	XT	-0.28 - 0.81	0.11	118.31 - 149.07	0.47 - 2.74	0.44 - 1.16	mc - bc
	AT	-0.28 - 0.82	-0.06	116.82 - 146.19	0.34 - 4.07	0.43 - 1.07	bd - bc
	PT	-0.06 - 0.80	0.22	119.15 - 147.17	0.68 - 1.93	0.46 - 1.10	pc - bc
	ON	-0.15 - 0.77	0.22	120.54 - 147.09	0.86 - 2.59	0.48 - 1.10	mc - bc
	OF	-0.25 - 0.76	0.31	121.61 - 144.12	0.63 - 4.13	0.50 - 1.02	bd - bc
	DR	-0.21 - 0.76	0.38	121.39 - 143.43	0.30 - 2.45	0.50 - 0.99	bd - bc
LA	**MT**	0.19 - 0.76	0.43	121.29 - 136.93	0.11 - 0.91	0.49 - 0.81	bd - xc
	XT	0.05 - 0.78	0.38	119.70 - 141.72	0.10 - 1.10	0.47 - 0.92	bc - xc
	AT	0.15 - 0.77	0.32	120.65 - 139.12	0.25 - 1.61	0.48 - 0.86	bc - xc
	PT	0.06 - 0.77	0.33	120.72 - 141.71	0.05 - 0.43	0.48 - 0.92	bc - xc
	ON	-0.12 - 0.81	0.62	117.38 - 143.89	0.17 - 1.57	0.43 - 1.00	bc - xc
	OF	-0.23 - 0.77	0.39	120.49 - 145.87	0.09 - 0.49	0.47 - 1.05	bc - xc
	DR	-0.16 - 0.81	0.49	117.70 - 144.77	0.13 - 1.03	0.43 - 1.02	bc - xc
LW NB	**MT**	0.35 - 0.60	0.53	126.03 - 140.14	0.73 - 1.32	0.58 - 0.88	pc - mc
	XT	-0.11 - 0.74	0.35	117.48 - 147.72	0.22 - 2.35	0.44 - 1.13	pc - xc
	AT	-0.85 - 0.72	0.42	118.30 - 156.33	0.65 - 5.21	0.45 - 1.47	pc - xc
	PT	-0.59 - 0.38	0.32	130.69 - 153.01	0.24 - 1.92	0.66 - 1.35	pc - xc
	ON	-0.69 - 0.39	0.15	133.22 - 154.70	1.12 - 2.06	0.73 - 1.40	pc - bd
	OF	-0.37 - 0.35	0.21	132.50 - 150.64	1.27 - 3.85	0.71 - 1.25	pc - bd
	DR	-0.51 - 0.38	0.18	131.31 - 152.61	0.86 - 2.96	0.70 - 1.32	pc - bd
LW SB	**MT**	-0.43 - 0.16	0.00	142.31 - 151.86	1.09 - 2.60	0.94 - 1.28	mc - bc
	XT	-0.12 - 0.25	0.09	141.08 - 147.43	1.35 - 3.58	0.91 - 1.10	xc - pc
	AT	-0.55 - 0.48	0.29	135.42 - 152.26	1.79 - 6.80	0.76 - 1.29	xc - pc
	PT	-0.55 - 0.23	0.00	142.24 - 152.09	2.29 - 4.48	0.94 - 1.28	xc - bd
	ON	-0.46 - 0.16	0.04	142.95 - 151.43	2.27 - 7.15	0.97 - 1.25	xc - bd
	OF	-0.48 - 0.10	-0.03	144.46 - 151.98	1.09 - 7.20	1.01 - 1.30	mc - pc
	DR	-0.46 - 0.13	-0.05	143.58 - 151.37	0.59 - 4.54	0.98 - 1.28	mc - pc
LW AB	**MT**	0.19 - 0.69	0.44	123.60 - 143.43	0.66 - 1.54	0.52 - 0.97	bd - mc
	XT	-0.14 - 0.74	0.23	120.33 - 148.43	0.63 - 3.46	0.47 - 1.15	pc - mc
	AT	-0.84 - 0.83	0.24	110.91 - 156.60	0.54 - 7.23	0.35 - 1.47	pc - xc
	PT	-0.72 - 0.64	0.15	125.02 - 154.99	0.47 - 2.47	0.56 - 1.42	pc - mc
	ON	-0.73 - 0.20	0.10	134.01 - 155.38	1.50 - 2.46	0.79 - 1.42	pc - mc
	OF	-0.30 - 0.30	0.08	132.94 - 149.65	1.75 - 4.59	0.75 - 1.21	pc - mc
	DR	-0.48 - 0.24	0.03	134.87 - 151.75	1.12 - 4.01	0.79 - 1.29	pc - mc
LE CB	**MT**	-0.12 - 0.58	0.37	74.99 - 129.73	0.02 - 0.50	0.11 - 0.63	mc - pc
	XT	-0.35 - 0.64	0.02	81.52 - 129.08	0.05 - 1.02	0.14 - 0.63	mc - pc
	AT	-0.25 - 0.33	0.02	78.74 - 135.00	0.04 - 0.67	0.13 - 0.77	mc - pc
	PT	-0.35 - 0.49	-0.10	87.60 - 129.75	0.03 - 1.31	0.17 - 0.66	bc - pc
	ON	-0.21 - 0.29	-0.05	82.38 - 136.33	0.02 - 0.56	0.15 - 0.80	bd - pc
	OF	-0.25 - 0.38	-0.13	82.55 - 135.17	0.04 - 0.96	0.15 - 0.75	mc - pc
	DR	-0.40 - 0.42	-0.10	87.42 - 133.11	0.05 - 0.77	0.17 - 0.72	mc - pc
LE EB	**MT**	-0.47 - 0.64	0.39	60.19 - 151.34	0.00 - 5.76	0.07 - 1.25	pc - mc
	XT	-0.11 - 0.63	0.15	60.83 - 147.00	0.00 - 3.86	0.07 - 1.09	pc - mc
	AT	-0.24 - 0.69	0.11	55.68 - 148.56	0.00 - 4.90	0.06 - 1.15	pc - mc
	PT	-0.30 - 0.59	0.05	62.12 - 149.23	0.00 - 3.53	0.08 - 1.17	bd - mc

	ON	-0.41 - 0.80	0.33	50.89 - 150.76	0.00 - 1.29	0.05 - 1.23	pc - mc
	OF	-0.51 - 0.65	0.13	59.58 - 152.10	0.00 - 1.78	0.07 - 1.27	pc - mc
	DR	-0.56 - 0.70	0.14	56.56 - 152.66	0.00 - 3.69	0.07 - 1.30	pc - mc
LE WB	MT	-0.66 - 0.47	-0.50	65.58 - 149.84	0.02 - 2.17	0.09 - 1.19	bd - bc
	XT	-0.43 - 0.30	-0.18	77.67 - 146.96	0.05 - 1.91	0.13 - 1.09	pc - xc
	AT	-0.47 - -0.17	-0.38	80.02 - 146.50	0.06 - 1.35	0.14 - 1.07	bc - bd
	PT	-0.73 - -0.44	-0.48	83.92 - 149.61	0.03 - 1.63	0.15 - 1.18	bc - mc
	ON	-0.59 - -0.06	-0.42	79.89 - 148.64	0.01 - 1.76	0.14 - 1.15	pc - xc
	OF	-0.58 - -0.27	-0.41	78.91 - 148.66	0.02 - 2.30	0.14 - 1.14	pc - mc
	DR	-0.66 - -0.32	-0.47	83.17 - 148.35	0.03 - 1.54	0.15 - 1.15	bc - mc
LE AB	MT	-0.20 - 0.59	0.30	75.96 - 134.05	0.03 - 0.76	0.12 - 0.72	bd - pc
	XT	-0.12 - 0.71	0.06	80.58 - 132.21	0.06 - 1.10	0.14 - 0.69	mc - pc
	AT	-0.21 - 0.61	-0.07	80.49 - 132.20	0.05 - 0.86	0.14 - 0.69	mc - pc
	PT	-0.38 - 0.63	0.10	86.98 - 127.54	0.04 - 1.41	0.17 - 0.60	bc - pc
	ON	-0.43 - 0.53	-0.05	83.69 - 136.86	0.03 - 0.93	0.15 - 0.79	bd - pc
	OF	-0.15 - 0.59	-0.01	82.35 - 130.25	0.04 - 1.07	0.15 - 0.65	bc - pc
	DR	-0.37 - 0.67	0.02	86.93 - 125.99	0.06 - 0.90	0.17 - 0.60	bc - pc
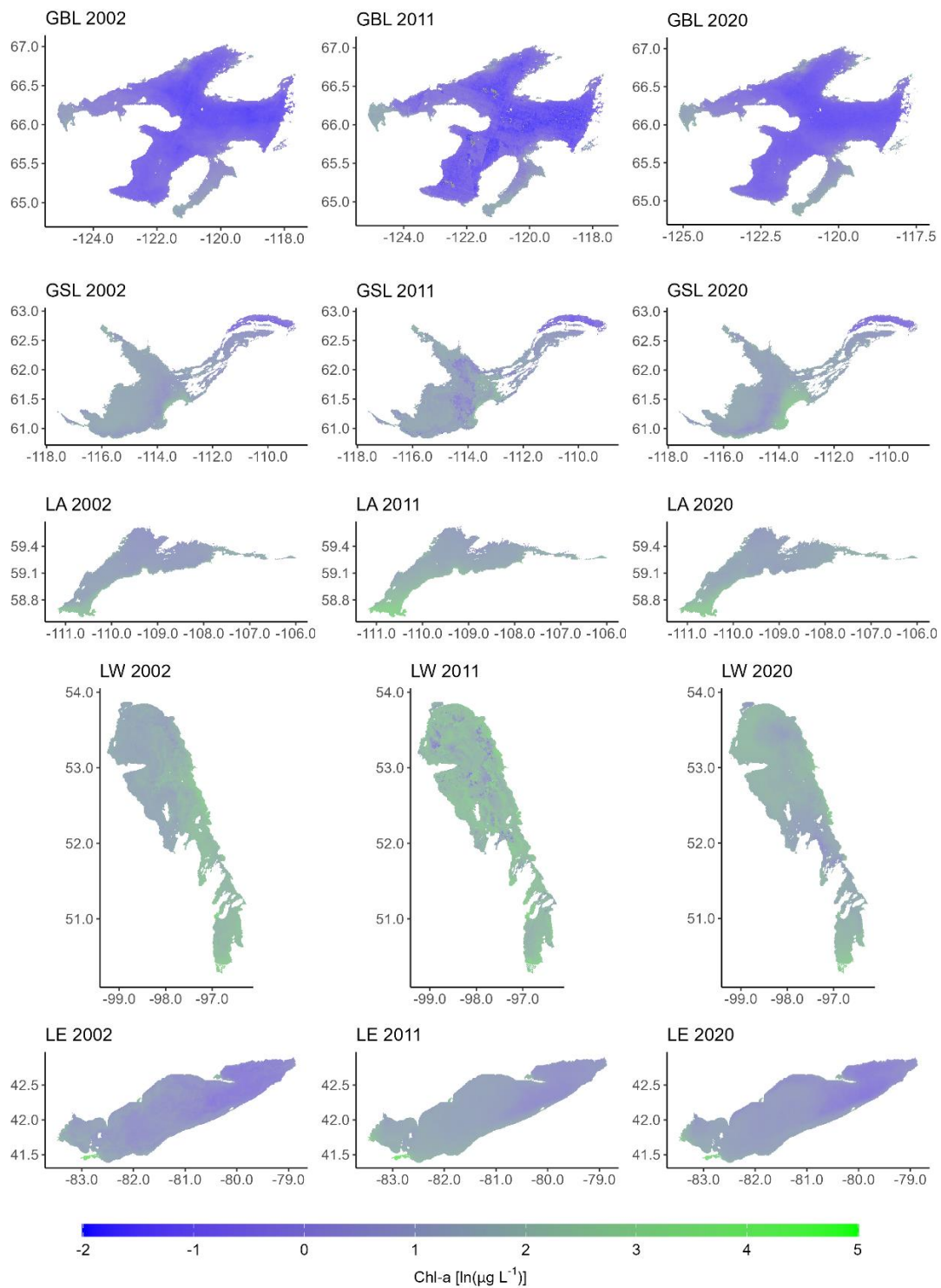
Figure B.7.1. Per pixel mean chl-*a* concentrations of three select years (2002, 2011, and 2020) for five North American Great Lakes where: GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie.

275
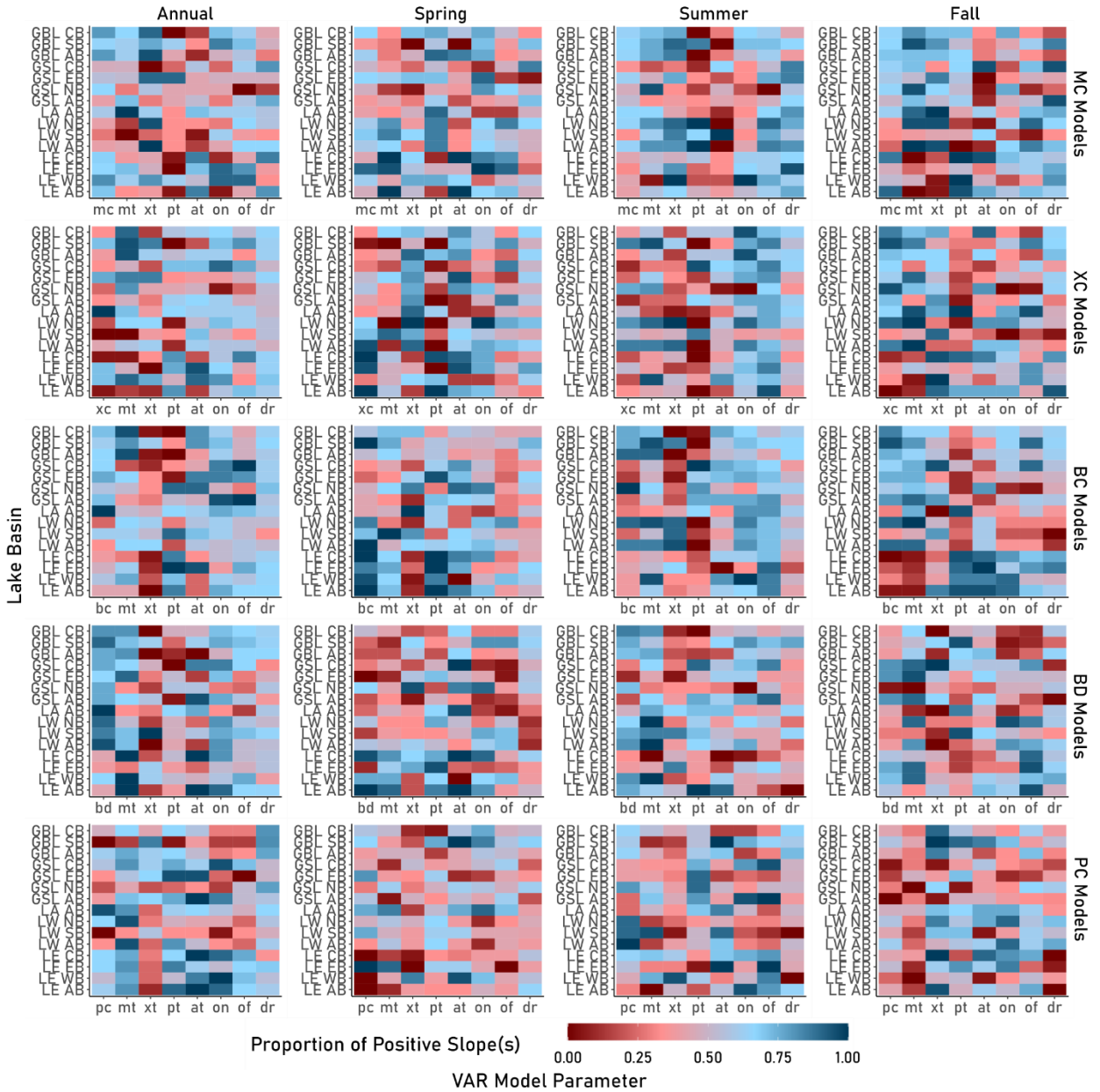
Figure B.7.2. The proportion of Vector Autoregression (VAR) slopes (β) which are positive for a given parameter. GBL = Great Bear Lake, GSL = Great Slave Lake, LA = Lake Athabasca, LW = Lake Winnipeg and LE = Lake Erie. AB = All Basins, CB = Central Basin, NB = North Basin, SB = South Basin, EB = East Basin, WB = West Basin. MC = Mean Chl-a, XC = Max Chl-a, BC = Abnormal Chl-a, BD = Abnormal Chl-a Days, PC = Peak Chl-a DOY, MT = Mean LSWT, XT = Max LSWT, AT = Anomaly LSWT Days, PT = Peak LSWT DOY, ON = Ice On DOY, OF = Ice Off DOY and DR = Ice Duration.