Chlorophyll-a Mapping in a Large Lake Using Remote Sensing Imagery: A Case Study of Western Lake Ontario

by

Ali Reza Shahvaran

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Earth Sciences (Water)

Waterloo, Ontario, Canada, 2024

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

Chapters 3 and 4 of this thesis each present a co-authored paper in manuscript format. As the first author, I was primarily responsible for the study design, execution, data collection, analysis, and writing. The following is the list of authors and their contribution for each chapter:

Chapter 3

Ali Reza Shahvaran (ARS), Homa Kheyrollah Pour (HKP), and Philippe Van Cappellen (PVC)

ARS took the lead in conducting the literature review, designing the methodology, collecting, and processing data, coding, analyzing, visualizing, and preparing the manuscript. HKP and PVC supervised the research, contributed to data collection, offered guidance on the methodology, workflow design, results, and discussion, provided feedback on visualizations, and participated in editing the manuscript.

Chapter 4

Ali Reza Shahvaran, Homa Kheyrollah Pour, Caren Binding (CB), and Philippe Van Cappellen

ARS conducted the literature review, data gathering, preparation, preprocessing, and processing, created visualizations, analyzed the results, and wrote the manuscript. HKP and PVC supervised the research, discussions, provided feedback on, and edited the manuscript. CB guided the methodology, workflow design, and contributed to the MDN model development, discussion, and manuscript review. All authors reviewed and approved the final manuscript.

Abstract

Western Lake Ontario (WLO) and Hamilton Harbour (HH) experience significant eutrophication challenges. Despite an overall decrease in the limiting nutrient phosphorus (P) inputs, recurrent nuisance (*Cladophora*) and cyanobacterial harmful algal blooms (cHABS) are observed in nearshore hotspots of WLO and HH, respectively. These events hint at a complex interplay of contributing factors including not only of P availability but nutrient enrichment in general, as well as invasive mussel species altering ecosystem dynamics, climate change, and other anthropogenic influences. As a result, continued and consistent monitoring is of paramount importance. Eutrophication in WLO and HH is also linked to the expanding urbanization within the Golden Horseshoe, which includes the Greater Toronto Area (GTA), along with nutrient point and nonpoint load sources from stormwater management systems and agricultural watersheds. Of importance are also the nutrient inputs flowing from Lake Erie through the Niagara River creating local productivity zones at the river mouth.

Traditional field-based monitoring methods face limitations, including high costs, labour intensity, limited temporal resolution and inadequate spatial coverage. In that respect, remote sensing (RS) may offer an alternative approach, leveraging the water colour (optical properties) to detect optically active constituents (OACs) like Chlorophyll-*a* (Chl-*a*) that can provide proxies for phytoplankton abundance in algae. The distinct spectral signatures of Chl-*a* make multi-spectral imagery a valuable tool for water quality assessment that can complement ongoing in-situ monitoring.

This thesis presents a comprehensive analysis aimed at enhancing the capacity for monitoring nearshore algal blooms in the oligo-mesotrophic WLO and eutrophic HH through publicly available high-spatial-resolution (< 100 m) RS satellites data, specifically Landsat 5, 7, 8, 9, and Sentinel-2. The research explores the optimal combinations of atmospheric correction methods and reflectance indexes to develop semi-empirical based Chl-*a* retrieval models specific to the (sub)regions considered. As an additional application, the satellite based Chl-*a* data are used to assess the spatial-temporal variability and trends of algal productivity over the past decade, identifying productivity hotspots and anomalies.

The thesis is structured in five chapters, beginning with a general introduction in Chapter 1, followed by Chapter 2, which offers the necessary background for understanding the research presented in the thesis. Chapters 3 and 4 delve into comparative evaluations of Chl-*a* retrieval methods and time-series analysis of algal bloom dynamics, respectively. The thesis ends with Chapter 5, which synthesizes the main findings and offers conclusions and future research directions.

Chapter 3 presents a comprehensive comparative evaluation of atmospheric correction processors and reflectance indexes, assessing their performance in Chl-*a* concentration retrieval from a multi-platform

collection of satellite data. By analyzing satellite scenes from different platforms alongside in-situ measured Chl-*a* data, the chapter develops predictive linear regression models. The results highlight the superior performance of certain combinations, particularly ACOLITE-corrected Landsat 8 and Sentinel-2 imagery utilizing two band ratio indexes, that is blue-to-green or blue-to-red, in capturing Chl-*a* concentration with acceptable accuracy.

Delving into the Chl-*a* dynamics, Chapter 4 presents a time-series analysis using Landsat 8 and 9 imagery from 2013 to 2023, to reconstruct the spatial-temporal patterns and hotspots in WLO and HH. After preprocessing a collection of Level-1 images with the optimal combination of atmospheric correction method and retrieval index, as identified in Chapter 3, a time-series collection of estimated Chl-*a* concentration maps are produced. By applying three algal growth indicators, namely bloom intensity, extent, and severity, along with averaging annual and monthly estimated Chl-*a* concertation maps and conducting a Mann-Kendall trend analysis, we are able to examine algal bloom dynamics, seasonality, and delineate areas of concern. The results should help in planning monitoring and design eutrophication management strategies for the region.

The findings from this thesis underscore the potential of space-borne RS in advancing water quality monitoring that can inform management practices. By identifying the most effective methods for Chl-*a* concentration retrieval and providing a nuanced understanding of algal growth dynamics, the research in this thesis contributes to both fields of aquatic RS and water quality monitoring. The comparative analyses, model developments, and spatial-temporal investigations not only offer practical tools for water quality assessment but also set the stage for future studies leveraging machine learning and existing satellite datasets. The work demonstrates the critical role of tailored RS applications in addressing eutrophication issues, advocating for integrated monitoring approaches to sustain aquatic ecosystems in the face of changing environmental conditions.

Acknowledgements

I extend my deepest gratitude to my supervisors, Dr. Homa Kheyrollah Pour and Dr. Philippe Van Cappellen, for their unwavering guidance and support throughout this research journey. Their expertise and encouragement have been pivotal to my success.

The support in funding for this thesis is from the Global Water Futures (GWF) program (Managing Urban Eutrophication Risks under Climate Change: An Integrated Modelling and Decision Support Framework), funded by the Canada First Research Excellence Fund (CFREF).

I also wish to thank my committee members, Dr. Helen Jarvie and Dr. Tonya DelSontro, for their insightful feedback and valuable time spent reviewing my thesis. Their contributions during committee meetings have greatly enriched my work.

Special acknowledgment goes to Dr. Caren Binding and Dr. Fereidoun Rezanezhad for their assistance crucial to this thesis development, and to Stephanie Slowinski for her resourcefulness both in the laboratory and the field.

My appreciation extends to Megan McCusker, Vanessa J. Beaulac, Dr. David Depew, and Carly Kemp, from Environment And Climate Change Canada (ECCC), and Hanna May and Kathleen Stevack from Ministry of the Environment, Conservation and Parks (MECP), for their support with data provision.

For the invaluable technical support, I am thankful to Gifty Attiah for assistance with code debugging, Dr. Bhaleka Persaud for data management, Grant Simpson and Mike Lackner for software support, and to Yuba Bhusal, Konrad Jens Krogstad, and Alex MacLean for their assistance in the field.

I am also grateful to Amir Reshadi, Dr. Jovana Radosavljevic, Dr. Amir Masoud Chegoonian, Dr. Zahra Akbarzadeh, and Dr. Michael Dallosch for sharing their insights and helping improve the quality of this work.

My heartfelt thanks to all members of the Ecohydrology Research Group (ERG) and the Remote Sensing of Environmental Changes (ReSEC) group, whose names have not been mentioned here, for their camaraderie and support.

Last but certainly not least, I must thank my family for their support, especially my wife, Elmira, without whose patience and sacrifice none of this would have been possible.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC	Atmospheric Correction
ACOLITE	Atmospheric Correction for OLI 'lite'
ACP	Atmospheric Correction Processor
AOC	Area of Concern
AOPs	Apparent Optical Properties
AW	Autumn/Winter
BOA	Bottom-of-Atmosphere
CDOM	Coloured Dissolved Organic Matter
cHABS	Cyanobacterial Harmful Algal Blooms
Chl-a	Chlorophyll-a
COG	Cloud-Optimized GeoTIFF
DEM	Digital Elevation Model
DN	Digital Number
DSF	Dark Spectrum Fitting
ECCC	Environment And Climate Change Canada
EH	Eutrophic/Hypereutrophic
EPA	Environmental Protection Agency
ESA	European Space Agency
EXP	Exponential Extrapolation
FAI	Floating Algae Index
GCPs	Ground Control Points
GEE	Google Earth Engine
GIS	Geographic Information System
GLNPO	Great Lakes National Program Office
GTA	Greater Toronto Area
HABs	Harmful Algal Blooms
HH	Hamilton Harbour
HLS	Harmonized Landsat and Sentinel-2
HPLC	High Pressure Liquid Chromatography
IOPs	Inherent Optical Properties
L1TP	Level-1 Terrain Corrected Product
LGL	Laurentian Great Lakes
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDAPE	Median Absolute Percentage Error
MDN	Mixture Density Network
MECP	Ministry of Environment, Conservation and Parks (Province of Ontario)
ML	Machine Learning
MSI	Multi-Spectral Instrument
NASA	National Aeronautics and Space Administration
NIR	Near-infrared

NOAA	National Oceanic and Atmospheric Administration
OACs	Optically Active Constituents
OCSMART	Ocean Color SMART (Specific Murky-water Algorithm for Retrievals and Turbidity)
OM	Oligotrophic/Mesotrophic
P	Phosphorus
RBG	Royal Botanical Gardens
RBINS	Royal Belgian Institute of Natural Sciences
RF	Random Forest
RGB	Red Green Blue
RMSE	Root Mean Squared Error
RMSLE	Root Mean Squared Logarithmic Error
RS	Remote Sensing
SABI	Surface Algal Bloom Index
SeaDAS	SeaWiFS Data Analysis System
SHAP	SHapley Additive exPlanations
SIAC	Spectral Index-based Atmospheric Correction
SNR	Signal-to-Noise Ratio
SS	Spring/Summer
TOA	Top-of-Atmosphere
TP	Total Phosphorus
TSI	Trophic State Index
TSS	Total Suspended Solids
UAV	Unmanned Aerial Vehicle
US	United States
USGS	United States Geological Survey
VIIRS	Visible Infrared Imaging Radiometer Suite
WLO	Western Lake Ontario

List of Symbols

		Unit
0^+, 0^-	Above and below water surface	-
$a(\lambda)$	Absorption at wavelength λ	nm^{-1}
$b_b(\lambda)$	Backward scattering at wavelength λ	nm^{-1}
$E_d(\lambda, 0^+)$	Downwelling irradiance just above the sea surface at wavelength λ	$W.m^{-2}.sr^{-1}.nm^{-1}$
$ESUN(\lambda)$	Mean solar exoatmospheric irradiance at wavelength λ	$W.m^{-2}.sr^{-1}.nm^{-1}$
$L(\lambda)$	Radiance at wavelength λ	$W.m^{-2}.sr^{-1}.nm^{-1}$
L_a	TOA radiance due to scattering by aerosols only	$W.m^{-2}.sr^{-1}.nm^{-1}$
L_{aR}	TOA radiance due to aerosol-molecule scattering	$W.m^{-2}.sr^{-1}.nm^{-1}$
L_R	TOA radiance due to molecular scattering in the atmosphere	$W.m^{-2}.sr^{-1}.nm^{-1}$
$L_{TOA}(\lambda)$ or L_t	Radiance measured at the top of the atmosphere by a satellite sensor at wavelength λ	$W.m^{-2}.sr^{-1}.nm^{-1}$
$L_W(\lambda)$	Water-leaving radiance at wavelength λ	$W.m^{-2}.sr^{-1}.nm^{-1}$
L_g^{TOA}	TOA radiance due to sunglint	$W.m^{-2}.sr^{-1}.nm^{-1}$
L_{sky}^{TOA}	TOA radiance due to surface-reflected background sky	$W.m^{-2}.sr^{-1}.nm^{-1}$
L_{WC}^{TOA}	TOA radiance due to whitecaps and foam	$W.m^{-2}.sr^{-1}.nm^{-1}$
p	p-value	-
r	Pearson's correlation coefficient	-
$R(\lambda)$ or $\rho(\lambda)$	Reflectance at wavelength λ	sr^{-1}
$R_{RS}(\lambda)$	Remote sensing reflectance (water colour) at wavelength λ	sr^{-1}
$R_S(\lambda)$ or $\rho_S(\lambda)$	Surface reflectance at wavelength λ	sr^{-1}
V	Cramér's V	-
θ or θ_s	Solar zenith angle	°
θ_v	Viewing direction	°
λ	Wavelength (i.e., satellite band)	nm
ρ	Spearman's correlation coefficient	-
ρ_{TOA} (λ)	Top-of-Atmosphere reflectance at wavelength λ	sr^{-1}
$\rho_w(\lambda)$	Water-Leaving radiance reflectance at wavelength λ	sr^{-1}
ϕ	Azimuthal angle measured relative to the Sun's azimuthal direction	°
χ^2	Chi-squared	-

Chapter 1
General Introduction

1.1 Motivation

The motivation for my research stems from the pressing issues of eutrophication in Hamilton Harbour (HH) and the nearshore zone of Western Lake Ontario (WLO), that experience nutrient enrichment due to urban and agricultural runoff, as well as other pressures from industrial activities, invasive species, and climate warming. These anthropogenic pressures have escalated the (re)occurrence of harmful (HABs) and nuisance algal blooms, degrading water quality, disrupting ecosystems, and imposing adverse socio-economic impacts on the region. Recognizing the limitations of traditional in-situ monitoring methods, motivated me to explore the potential of space-borne optical remote sensing (RS) as a complementary tool, presenting a promising avenue for historical and near-real-time water quality assessment.

Driven by the need to refine remote sensing methodologies for monitoring algal abundance, Chapter 3 evaluates various paths for producing case-specific RS-derived Chlorophyll-*a* (Chl-*a*) concentration models to identify the most accurate ones. This effort underscores the importance of selecting an appropriate scheme – a combination of atmospheric correction processor and Chl-*a* retrieval index – which directly impacts the precision of remote sensing analyses, especially in oligo-mesotrophic conditions where retrieving low Chl-*a* concentration is challenging.

Having tested different retrieval approaches, Chapter 4 builds on this foundation by selecting the proper scheme for conducting a time-series analyses of algal abundance in WLO and HH. The chapter shows the potential of space-borne RS in filling the gaps of in-situ data during certain months (especially during winter) and years (e.g., 2020 due to COVID), by providing a comprehensive mapping of algal bloom dynamics and hotspots over the last eleven years. Additionally, my work helps alleviate the lack of high-spatial-resolution satellite-derived studies on Lake Ontario's waters, in contrast to the well-studied eutrophic Lake Erie.

I hope my findings will serve as a valuable resource for both local and global stakeholders, enhancing decision-making capabilities for local authorities in managing water quality and providing an integrated, multidimensional insight into algal dynamics, including bloom occurrences, in the region from 2013 to 2023, as well as contributing to advancing the precision and effectiveness of Chl-*a* retrieval methodologies through the use of space-borne data at a global scale.

1.2 Research Objectives

The overall objective of this thesis is to enhance the effectiveness and application of space-borne multi-spectral remote sensing to map the distributions of Chl-*a* – in space and time. As a case study, I focus on the western basin of Lake Ontario (WLO) and Hamilton Harbour, an embayment connected to WLO by a shipping channel. The working hypothesis is that by refining RS-based water quality assessments existing gaps in eutrophication mapping can be overcome. The specific objectives of each chapter are as follows:

Chapter 3 objectives:

- Evaluating the performance of 27 Chl-*a* retrieval indexes across eleven Level-2 products from four satellites (Landsat 5, 7, 8 and Sentinel-2), using 600 in-situ matchups collected between 2000-2022.
- Developing Chl-*a* retrieval regression models in order to compare their performance across 28 scenarios, depending on data categorization and satellite platform.

Chapter 4 objectives:

- Producing Chl-*a* maps from 2013-2023 with Landsat 8 and 9 imagery for detailed spatial-temporal analysis by employing three algal growth indicators – intensity, extent, and severity – and Geographic Information System (GIS) products such as averaged annual and monthly estimated Chl-*a* concentration, as well as trend maps to observe spatial-temporal variability, trends, and identify (re)emerging productivity hotspots.
- Validating the findings with in-situ data as well as published reports from other sources.
- Offering insights into algal blooms at locations and times with in-situ data gaps (for instance, during COVID-caused shutdowns), and suggesting monitoring strategies informed by the identified seasonal trends and hotspots.

1.3 Research Contributions and Impacts

Each of the core chapters (i.e., 3 and 4) of this thesis makes a unique contribution to the literature of aquatic remote sensing and offers distinct applications for water quality monitoring. Chapter 3, submitted to the journal Remote Sensing and currently under review, underscores the importance of the Chl-*a* retrieval indexes using various sensors and highlights the critical role of satellite selection, in-situ data categorization, and atmospheric correction in enhancing model accuracy. The breadth of this

analysis, encompassing 600 matchup data points, the employment of four satellite platforms, and the application of eleven atmospheric corrections alongside 27 retrieval indexes, sets this chapter apart from similar studies in terms of scope and extent. Furthermore, the comprehensiveness of matchup data, in terms of counts and range, primarily within the oligo-mesotrophic range, where performance assessment is notably more challenging than in eutrophic conditions, further enhances the chapter's contribution to the field. The findings of this chapter can be used, with caution, in case studies with similar optical conditions, for methodology development, particularly in selecting the appropriate atmospheric correction for preprocessing and applying the most sensitive reflectance index for retrieving Chl-*a* concentration, based on each study's data range and satellite platform used.

Chapter 4 overcomes the limitations of conventional field surveys in algal growth monitoring by utilizing Landsat 8 and 9 imagery to reveal the spatial-temporal dynamics of algal blooms from 2013 to 2023. The insights gained from this chapter not only deepen the understanding of algal dynamics in the region through satellite-derived maps and indicators but also enable the efficient allocation of monitoring resources, such as timing lake sampling to coincide with bloom periods or increasing sampling frequency in Chl-*a* hotspots. The outcomes of this chapter are anticipated to benefit environmental agencies in decision-making regarding monitoring programs and the mitigation of algal blooms in WLO and HH.

1.4 Thesis Outline

This thesis comprises five chapters, structured to systematically address the analysis of Chl-*a* retrieval methods and their application to map Chl-*a* concentration time-series in WLO and HH. Chapter 1 sets the stage with an introduction, outlining the motivation, research objectives, contributions, impacts, and the thesis outline. Chapter 2 delves into the necessary background and theory, covering water colour remote sensing, the importance of atmospheric correction, and semi-empirical retrieval of Chl-*a* concentration, culminating with a summary. Chapters 3 and 4, presented in manuscript format, explore diverse aspects of remote sensing: Chapter 3 compares semi-empirical retrieval schemes for Chl-*a* concentration, and Chapter 4 conducts a time-series analysis of RS-derived Chl-*a* concentration using Landsat 8 and 9 imagery. Chapter 3 and 4 follow a consistent structure of: Introduction, Materials and Methods, Results, Discussion, and Conclusion. Chapter 5 concludes with a summary of key findings, remaining research challenges, potential applications, and recommendations for future research. Supplementary to the main text, the thesis includes references, a glossary defining technical terms and jargon, and two appendices providing additional materials for Chapters 3, and 4, respectively.

Chapter 2
Background and Theory

2.1 Introduction

The assessment and monitoring of water quality are crucial for the protection and sustainable use of inland waters such as lakes, which play a pivotal role in environmental systems (Sagan et al., 2020). Serving as habitats for diverse species and being integral to nutrient and carbon cycles, these water bodies are essential components of our natural world (Dörnhöfer & Oppelt, 2016). However, they face significant threats from anthropogenic activities and climate change, leading to degradation through contamination with pollutants like heavy metals, nutrients, and microorganisms (Dörnhöfer & Oppelt, 2016). Traditional monitoring methods, primarily field-based, have been the mainstay but are often hampered by their labour-intensive nature, high costs, and inability to effectively capture the temporal and spatial variability of water quality phenomena (Dube et al., 2023). Over the past few decades, satellite remote sensing technologies have emerged as vital, complementary tools to these traditional methods, offering a means to retrieve and analyze water quality proxies over vast spatial and temporal scales (Tavora et al., 2023). By measuring the water-leaving radiance in the visible to near-infrared spectrum, remote sensing can infer properties related to water clarity, biota, and hydrology, thus providing essential information for understanding the dynamics of various water quality indicators, such as Chl-*a*, and other optically active constituents (OACs) (Dörnhöfer & Oppelt, 2016).

Remote sensing operates on the principle of analyzing the radiation that is either reflected off the surface of water bodies or emitted (thermal) by them. This process involves the consideration of spatial, temporal, spectral and radiometric resolutions of the sensors used. Spatial resolution defines the smallest area on the ground that a single pixel in the image represents, which can range from tens of meters in Landsat images to centimetric resolutions in unmanned aerial vehicle (UAV)-mounted sensors, for example. Temporal resolution indicates the frequency with which a sensor captures images of the same area, varying from once a day to once every few weeks, depending on the orbiting platform. Spectral resolution refers to the number of spectral bands or wavelengths of light that the sensor can measure, impacting the ability to discern different materials or phenomena on the Earth's surface. Last, but not least, radiometric resolution is the sensitivity of a sensor to detect subtle differences in the intensity of electromagnetic energy, enabling it to distinguish between different levels (shades) of brightness (in a certain wavelength).

The evolution of high-spatial-resolution (< 100 m) remote sensing over the last 50 years, particularly through the Landsat series of satellites, has enhanced the ability to study and monitor the water quality of coastal and inland waters (Tavora et al., 2023). These technologies have been instrumental in identifying and quantifying the distribution of key water quality parameters, albeit with the necessity of performing atmospheric correction to mitigate the influence of atmospheric scattering and absorption on the captured signals. The integration of data from various sensors, including Landsat 8 and 9 and Sentinel-2, alongside in-situ measurements, has facilitated the development of models for estimating Chl-*a* concentration and other indicators at regional to global scales, underscoring the importance of remote sensing in proactive water quality monitoring.

2.2 Water Colour Remote Sensing

Water colour remote sensing is the measurement of the colour of the water from satellites or aircraft to determine water properties and constituents (see Figure 2.1). The essence of water colour remote sensing lies in deciphering the intricate relationship between the concentration of water constituents and the scattering signals – known as water-leaving radiance – captured by these sensors (H. Yang et al., 2022). For OACs, such as Chl-*a*, Total Suspended Solids (TSS), and coloured dissolved organic matter (CDOM) (see Figure 2.2), models can be established based on the relationship between inherent optical properties (IOPs) and remote sensing reflectance.

The interaction of sunlight with the Earth's surface, particularly with water bodies, is a complex process influenced by multiple factors. As sunlight passes through the atmosphere, it can either be reflected off the water surface or penetrate the water, where it is then either absorbed or scattered by various OACs, including phytoplankton, non-algal particles, and CDOM (Dörnhöfer & Oppelt, 2016). The fate of absorbed light and its scattering direction – whether forward or backward – determines the radiance that is eventually detected by sensors. In the context of aquatic remote sensing, the focus is on the remote sensing reflectance (R_{RS}), a critical metric in water colour remote sensing that is defined by the ratio of backscattering to the total absorption and backscattering of light as in:

$$R_{RS}(\lambda, 0^+) \cong C\frac{b_b(\lambda)}{a(\lambda)+b_b(\lambda)} = \frac{L_w(\lambda)}{E_d(\lambda,0^+)} \quad (2.1)$$

where a is absorption by OACs, b_b is the backward scattering, L_w is the water leaving radiance, E_d is downwelling irradiance and R_{RS} is the remote sensing reflectance.

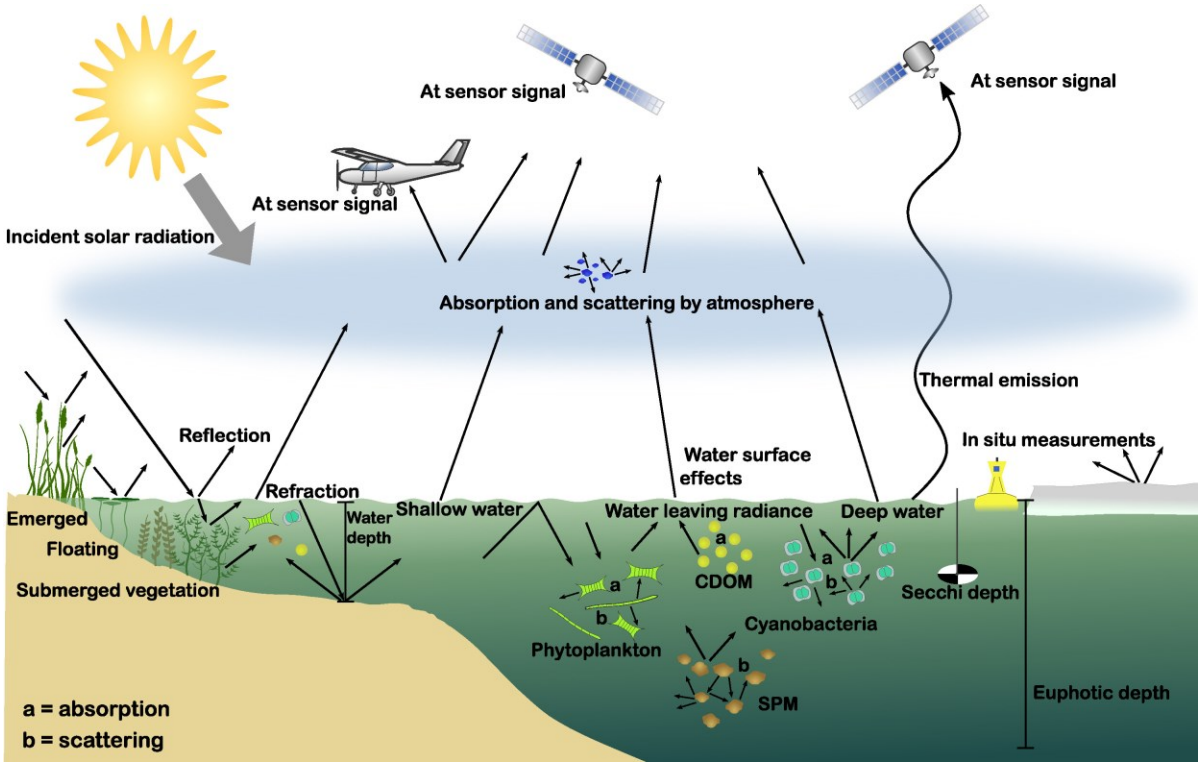

Figure 2.1. Interaction between water colour remote sensing indicators from Dörnhöfer & Oppelt (2016) (reuse permission acquired through Elsevier and Copyright Clearance Center)

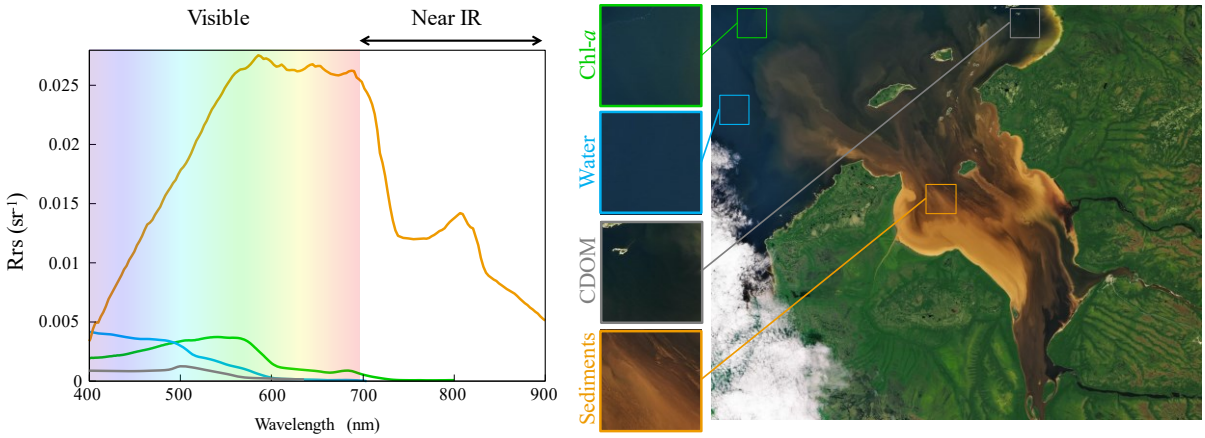

Figure 2.2. Visualization of four optically active constituents (OACs) displayed on a true-colour Landsat 8 image on the right, alongside their schematic spectral signatures in the visible and near-infrared (NIR) range on the left, adopted from Mehta & Flores (2018) with reuse permission.

2.3 Importance of Atmospheric Correction

Atmospheric correction is a critical process in the analysis of satellite imagery, particularly in the field of water colour remote sensing. The radiance measured at the top-of-atmosphere (TOA) by satellite sensors is a composite signal, comprising radiance emanating from within the water, radiance interacting with the water surface, and radiance scattered by atmospheric gases and aerosols all effectively summarizes in equation 2.2 (Tavora et al., 2023):

$$L_t = L_R + [L_a + L_{aR}] + L_g^{TOA} + L_{sky}^{TOA} + L_{wc}^{TOA} + L_w^{TOA} \quad (2.2)$$

where each component of the total radiance at TOA (L_t) represents different sources of scattering and absorption, from molecular (L_R) and aerosol (L_a and L_{aR}) scattering in the atmosphere to specific contributions from sunglint (L_g^{TOA}), surface-reflected background sky (L_{sky}^{TOA}), whitecaps and foam (L_{wc}^{TOA}), and the critical water-leaving radiance (L_w^{TOA}) all measured in the unit of radiance W m^{-2} sr^{-1} nm^{-1}. The significance of atmospheric correction becomes evident when considering that 90% of the signal captured by remote sensors is directly affected by the atmosphere, leaving only a small fraction (10%) as the water-leaving radiance of interest for remote sensing analyses (Dörnhöfer et al., 2018; Tavora et al., 2023). This overwhelming influence of atmospheric necessitates the accurate removal of these contributions to isolate the pure signal emanating from water bodies.

The complexity of atmospheric correction is compounded by the diverse optical properties of atmospheric particles and gases, which require sophisticated algorithms to correctly remove their effects from the satellite-detected signal. Different water bodies, classified as case one (primarily phytoplankton) and case two (varied concentrations of optically active substances), demand distinct approaches to atmospheric correction due to their unique absorption and scattering characteristics. Traditional atmospheric correction algorithms may over- or under-correct the signal in case two waters, highlighting the need for specialized processors tailored to the specific conditions of each water body. The development and application of various aquatic atmospheric correction processors, such as ACOLITE, POLYMER, and C2RCC, reflect the ongoing efforts to refine this process, offering more reliable methods to derive water-leaving reflectance and, subsequently, water quality parameters from satellite imagery.

The success of atmospheric correction is crucial not only for the accurate retrieval of water-leaving radiance but also for the subsequent derivation of water quality products. Given the large atmospheric contribution to the measured TOA radiance, the choice of atmospheric correction processor and the algorithms used to derive products from above-water reflectance can significantly impact the estimates

of these water quality parameters. As showcased in Figure 2.3, the selection of atmospheric correction can tremendously affect the derived R_{RS} especially in the blue to green range.

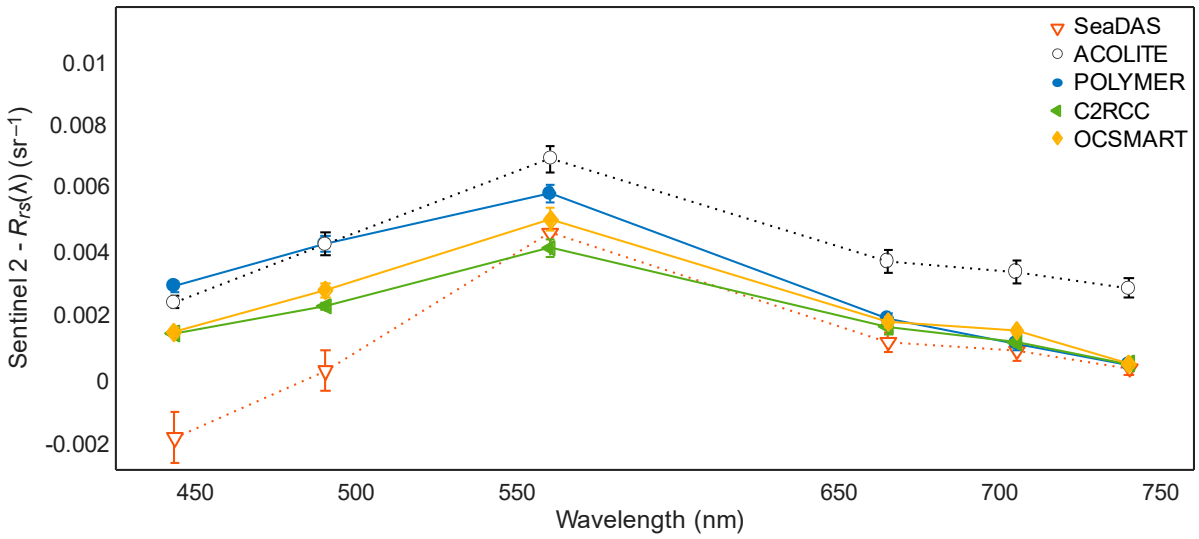

Figure 2.3. Comparing the performance of various atmospheric correction methods for deriving $\boldsymbol{R_{RS}}$ from a Sentinel-2 image, adopted from Tavora et al. (2023), reused under the CC BY 4.0 license.

2.4 Semi-Empirical Retrieval of Chl-*a* Concentration in Water

The retrieval of Chl-*a* concentration from aquatic environments using remote sensing technologies has evolved through the application of various bio-optical models, which can be broadly categorized as empirical, semi-empirical, semi-analytical, quasi-analytical, and analytical (D. R. Mishra et al., 2017). Each of these methodologies approaches the challenge of quantifying water quality parameters with differing degrees of reliance on direct measurements, statistical correlations, and physical understandings of light interactions within water bodies (Dörnhöfer & Oppelt, 2016; Gholizadeh et al., 2016; D. R. Mishra et al., 2017). Empirical models establish a direct statistical relationship between in-situ measurements of Chl-*a* and remotely sensed reflectance, typically requiring extensive site-specific calibration. In contrast, semi-analytical and quasi-analytical models derive from the inversion of radiative transfer equations, attempting to directly link apparent optical properties (AOPs) like reflectance, with IOPs of water constituents (D. R. Mishra et al., 2017).

Unlike purely empirical models that rely on statistical relationships without physical grounding, semi-empirical models integrate physical principles with empirical data, enabling more versatile applications across different water types and conditions (D. R. Mishra et al., 2017). Semi-empirical models,

particularly relevant for Chl-*a* retrieval, capitalize on specific spectral features of water constituents that influence light absorption and scattering. The successful application of semi-empirical models for Chl-*a* retrieval is underscored by their ability to interpret the spectral signatures of waters with varying Chl-*a* concentration. For instance, waters with high Chl-*a* concentration exhibit distinct spectral characteristics, such as local extrema at 550 nm and 685 nm, reflecting the direct influence of pigments on the light absorption and scattering. Yet, the bands of 685 to 550 nm are not the sole bands for estimating Chl-*a* concentration; there exist numerous numerical retrieval indexes based on a variety of bands, that are tailored for this purpose (H. Yang et al., 2022). For example, a commonly used index in the analysis of VIIRS images is the ratio of 550 to 480 nm. As illustrated schematically in Figure 2.**4**, the remote sensing reflectance for different concentrations of Chl-*a* demonstrates how the ratio of 550 to 480 nm can be instrumental in estimating Chl-*a* concentration effectively. This method, refined through validations against in-situ measurements, exemplifies the practical utility, simplicity and most importantly reproducibility of semi-empirical models in translating spectral data into meaningful estimations of water quality parameters.

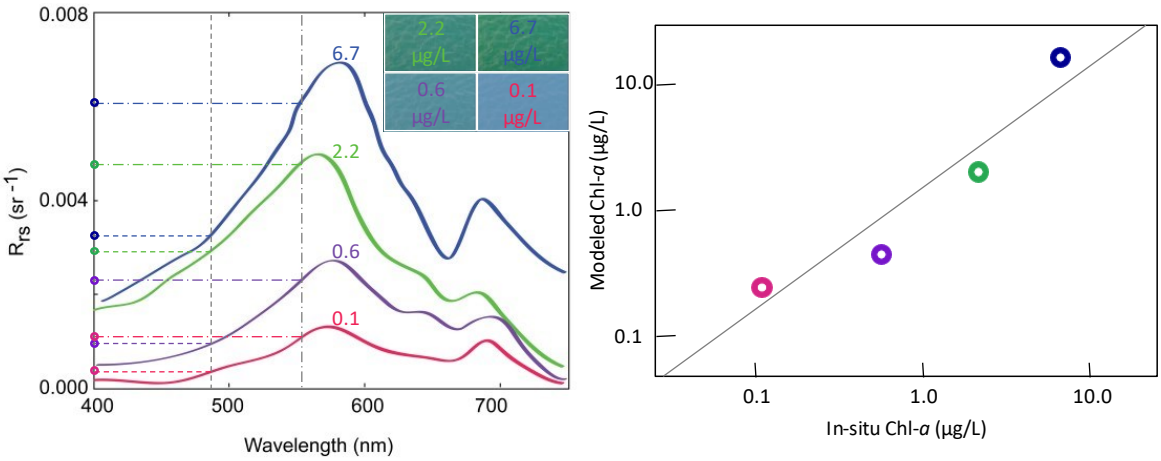

Figure 2.4. Schematic representation of remote sensing reflectance across different concentrations of Chl-*a*, highlighting the utility of the 550 nm to 480 nm ratio for Chl-*a* estimation in VIIRS images, adopted from Mehta & Flores (2018) with reuse permission.

2.5 Summary

Chapter 2 serves as the foundation for this thesis by reviewing the principles and methodologies used for the remote sensing of Chl-*a* concentration in oligo-mesotrophic waters of Western Lake Ontario and Hamilton Harbour. The chapter underlines the importance of assessing and monitoring water

quality, then delves into the nuances of remote sensing, detailing the critical resolutions – spatial, temporal, spectral, and radiometric – that enable the accurate capture and analysis of water quality parameters over extensive areas and periods. This discussion sets the stage for a deeper exploration of water colour remote sensing and the significance of atmospheric correction and outlines the underpinnings and the diverse modeling approaches – ranging from empirical to semi-empirical and analytical methodologies – employed in the retrieval of Chl-*a* with an emphasis on semi-empirical models that are being used in the subsequent two chapters.

Chapter 3
Comparative Evaluation of Semi-Empirical Approaches for Satellite-Derived Chlorophyll-*a* Concentration Retrieval in the Nearshore and Offshore Waters of a Large Lake (Lake Ontario)

3.1 Introduction

Eutrophication of freshwater bodies is a global environmental concern. It is driven primarily by nutrient enrichment, most often accelerated by anthropogenic activities such as agricultural fertilizer application, wastewater discharge, and urbanization (Mamun et al., 2021). Cultural eutrophication promotes excessive growth of algae, leading to deterioration of water quality and overall ecosystem health, and the loss of aquatic habitats (Li et al., 2021). In extreme cases, harmful algal blooms (HABs) occur through the rapid proliferation of algae, in particular cyanobacteria (Sagan et al., 2020). The occurrence of HABs is influenced by a complex interplay of biogeochemical and physical factors, including nutrient availability (mainly phosphorous and nitrogen), water temperature, and hydrodynamics (Adams et al., 2022).

In the western portion of Lake Ontario algal blooms are a particular concern in nearshore areas including Hamilton Harbour (HH), the Toronto shoreline, and the Niagara River outlet (Markovic et al., 2019). The major drivers of eutrophication in these areas are excess nutrient runoff associated with rapid urbanization of the Greater Toronto Area (Higgins et al., 2012) and surrounding agricultural and industrial activities (Hui et al., 2021). Other contributing factors include nutrient-rich inflow from Lake Erie via the connecting Niagara River (Malkin et al., 2010), changes in water temperature and circulation patterns (Blagrave et al., 2022), and ecological disturbances such as the invasion by nuisance mussels (Blagrave et al., 2022). Given these rapidly changing dynamics, continued water quality monitoring of the nearshore and offshore waters of Western Lake Ontario (WLO) remains essential.

Traditional field-based water quality monitoring methods have played an integral role in tracking and assessing trophic changes in large lakes over the past several decades. While these methods are essential, they face limitations due to their high costs and personnel requirements, as well as their restricted spatial and temporal coverage. In this context, remote sensing (RS) techniques have emerged as a valuable complimentary tool, enhancing traditional water quality monitoring with their ability to provide large-scale, frequent, and cost-effective observations of water bodies (Gholizadeh et al., 2016). Multispectral satellites like Landsat 5, 7, 8, and Sentinel-2 are now extensively used for lake eutrophication monitoring at a global scale (Beck et al., 2016). In RS research, Chlorophyll-*a* (Chl-*a*)

concentration, as the key photosynthetic pigment, is most commonly used to map algal abundance in surface waters and monitoring eutrophication trends (Maeda et al., 2019; Pirasteh et al., 2020). The accuracy of Chl-*a* estimations highly depends on the applied atmospheric correction and the selection of appropriate retrieval algorithms (Grendaitė & Stonevičius, 2022). By carefully addressing these aspects, RS can provide essential data on water quality trends and dynamics.

The estimation of Chl-*a* concentration using multispectral satellite data has been growing in recent years (H. Yang et al., 2022). This trend is notably tied to the launch of the Sentinel-2 satellite constellation in 2015, which added a red-edge band and increased spatial resolution down to 10 meters. High spatial resolution is particularly important for monitoring nearshore waters where land adjacency is a challenge. Further improvements are expected in the near future with the anticipated launch of NASA's 10-m super-spectral Landsat Next in 2030, which promises even finer spectral resolution, likely improving water colour remote sensing. A limited number of recent studies in the 10- to 30-meter resolution range have compared the performance of different atmospheric corrections and Chl-*a* concentration retrieval indexes (Table 3.1). Not surprisingly, these studies show that models to retrieve Chl-*a* concentration from satellite data tend to perform best for the data range and the region they were trained on.

Our work builds on the previous studies in Table 3.1 by including expanded ranges of Chl-*a* retrieval indexes, atmospheric correction processors, satellites, and in-situ data. Specifically, we evaluate the performance of 27 of the most used Chl-*a* concentration retrieval indexes applied to 11 types of Level-2 products from four distinct sensors, paired to 600 near-synchronous, co-located in-situ matchups in the western basin of Lake Ontario. We focus on Lake Ontario because, unlike upstream Lake Erie, whole-lake assessments of spatial and temporal trends in algal abundance remain limited. We aim to contribute to filling this gap using 10- and 30-meter resolution satellite imagery applied to the western part of the lake.

The Chl-*a* data are categorized according to seasonality, location, and Trophic State Index (TSI). Each data category is then calibrated for a given sensor using the in-situ matchup data. The results form the basis of a comparative analysis of the sensors' performances and the effectiveness of the Chl-*a* concentration estimation schemes for both nearshore and offshore waters of the lake. Our work should help support future RS studies on the water quality and ecosystem health of Lake Ontario and those of large lakes in general.

Table 3.1. Summary of recently published studies comparing atmospheric corrections and Chl-*a* retrieval indexes

Study Reference	Number of Chl-*a* Retrieval Indexes	Water-Truthing*		Imagery Matchups			Temporal Coverage of Matchups
		Number of Chl-*a* Data Points	Radiometric Matchup Availability	Comparable** Sensors	Number of Scenes	Atmospheric Corrections	
(Pahlevan et al., 2021)	5	123	✓	OLI, MSI	N/A	ACOLITE, GRS, MEETC2, Polymer, SeaDAS, iCOR	N/A
(Abdelal et al., 2022)	8	~50	✓	OLI, MSI	3	DOS, ATCOR, DSF, EXP, L8SR	2019
(Warren et al., 2019)	2	N/A	✓	MSI	~60	ACOLITE, C2RCC, iCOR, l2gen, Polymer, Sen2Cor	2016 – 2017
(Boucher et al., 2018)	6	351	✗	OLI	11	DOS	2013 – 2015
(Sòria-Perpinyà et al., 2022)	9	146	✓	MSI	44	C2RCC, C2X, C2XC, and Polymer	2017 – 2021
(Tavares et al., 2021)	6	97	✓	MSI	N/A	ACOLITE, C2RCC, GRS, iCOR, SeaDAS, Sen2Cor	2017 – 2019
(Saberioon et al., 2020)	19	~40	✗	MSI	10	ACOLITE	2017 – 2018
(Tian et al., 2022)	4	139	✓	OLI	61	SeaDAS, ACOLITE, C2RCC, iCOR	2019 – 2021
(Nazeer & Nichol, 2016)	17	120	✗	TM, ETM+	27	6S	2000 – 2012
(Buma & Lee, 2020)	4	40	✗	OLI, MSI	3	FLAASH, QUAC	2015
(Soriano-González et al., 2019)	7	27	✗	MSI	13	ACOLITE	2016 – 2018
(Barreneche et al., 2023)	17	214	✗	OLI, MSI	92	SeaDAS, POLYMER, ACOLITE	2004 – 2020
(Deutsch et al., 2018)	3	108	✗	ETM+, OLI	16	LEDAPS, L8SR	2013 – 2015
(Ansper & Alikas, 2019)	28	12	✓	MSI	N/A	ACOLITE, C2RCC, POLYMER, Sen2Cor	2015 – 2017
(Ha et al., 2017)	9	30	✓	MSI	2	ELM	2016 – 2017
(Rodríguez-López et al., 2020)	11	39	✗	TM, ETM+, OLI	14	DOS	2001 – 2003 2017 – 2019
(Ogashawara et al., 2021)	5	32	✓	MSI	1	ACOLITE, C2RCC, C2X, iCOR, MAIN, Sen2Cor	2019
(Soomets et al., 2020)	21	41	✓	MSI	41	C2RCC, C2X	2018
(Grendaitė & Stonevičius, 2022)	10	30	✓	MSI	7	ACOLITE, iCOR, Sen2Cor, C2RCC, C2X, POLYMER	2018 – 2019
Current Study	27	600	✗	TM, ETM+, OLI, MSI	236	LEDAPS, LaSRC, Sen2Cor, ACOLITE, ATCOR, C2RCC, DOS 1, FLAASH, iCOR, Polymer, QUAC	2000 – 2022

* The word "truth" although common, might be misleading, as field data measurement inherently involves error.

** Only limited to the list of sensors used in this study. Also, studies based on simulated satellite imagery are excluded from the table.

Abbreviations in alphabetical order: 6S: Second Simulation of the Satellite Signal in the Solar Spectrum, ACOLITE: Atmospheric Correction for OLI Lite, ATCOR: Atmospheric and Topographic Correction, C2RCC: Case 2 Regional CoastColour, C2X: Case 2 eXtreme, C2X-COMPLEX, DOS: Dark Object Subtraction, DSF: Dark Spectrum Fitting, ELM: Empirical Line Method, ETM+: Enhanced Thematic Mapper Plus, FLAASH: Fast Line-of-sight Atmospheric Analysis of Hypercubes, GRS: Glint Remove Sentinel, LaSRC: Landsat Surface Reflectance Code, L8SR: Landsat 8 OLI Surface Reflectance, LEDAPS: Landsat Ecosystem Disturbance Adaptive Processing System, MAIN: Modified Atmospheric Correction for Inland Waters, MSI: Multispectral Instrument, OLI: Operational Land Imager, QUAC: Quick Atmospheric Correction, SeaDAS: SeaWiFS Data Analysis System, Sen2Cor: Sentinel-2 Atmospheric Correction, TM: Thematic Mapper

3.2 Materials and Methods

3.2.1 Study Site

Lake Ontario, the smallest and most easterly of the Great Lakes, ranks as the 13th largest lake in the world (Huang et al., 2012). It has a drainage area of around 64,000 km^2, a surface area of 19,000 km^2, and a total volume of 1,650 km^3. The lake's mean depth is 87 m, and the maximum depth is 244 m. The water residence time is approximately 6-8 years, and the lake's shoreline extends over 1,150 km. Our study focused on the western basin of the lake, abbreviated WLO, which represents a critical water resource for an estimated nine million people (Blagrave et al., 2022). WLO covers about one third of the entire surface area of Lake Ontario (Figure 3.1). Also included in our study is Hamilton Harbour (HH), a 20 km^2 embayment at the western tip of WLO connected to the lake via a shipping channel. Since 1987, HH has been designated an Area of Concern (AOC) under the Canada-US Great Lakes Water Quality Agreement. As a result, HH has a much higher density of water quality monitoring stations compared to WLO as can be seen in Figure 3.1.

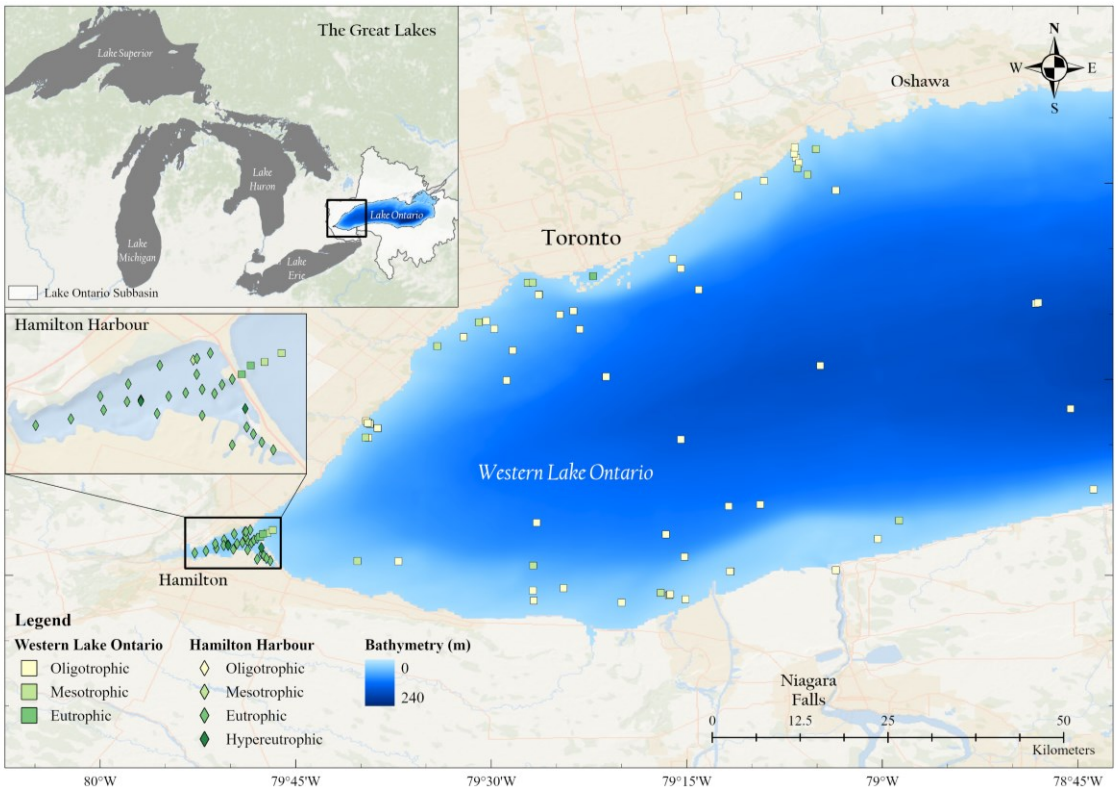

Figure 3.1. Map of the study area showing in-situ measurement locations of matchup data. Diamond and square markers represent Hamilton Harbour (HH) and Western Lake Ontario (WLO) measurements, respectively. Each marker is colour-coded according to the respective trophic state.

Historically, WLO has experienced water quality issues due to both urban and rural non-point sources of nutrients and contaminants delivered to the lake via rivers and storm sewers (Munawar & Fitzpatrick, 2018). The lake features mesotrophic to eutrophic nearshore zones and an oligotrophic offshore core (E. Howell, 2018). In recent years, the total phosphorus (TP) concentration in the lake ranges on average from 6 to 8 µg/L, while the Chl-*a* concentration varies between 0 and 17 µg/L, depending on the location (Auer et al., 2021; Dove & Chapra, 2015). As expected, HH exhibits significantly higher TP and Chl-*a* concentration, with the latter often exceeding 10 µg/L, that is the threshold value commonly used to define an algal bloom (Binding et al., 2021).

The primary drivers of eutrophication in WLO and HH include rapid urbanization, ecosystem and climate change (Higgins et al., 2012; E. Howell, 2018; Hui et al., 2021), while invasive dreissenid mussels transform phosphorus into more readily bioavailable forms that promote algae growth (Higgins et al., 2012; E. Howell, 2018; Mohamed et al., 2019). The onset, duration, and intensity of algal growth are further influenced by environmental factors such as temperature, sunlight, and wind mixing (Adams et al., 2022; Binding et al., 2021). The harmful effects of eutrophication in the region include water supply issues, human and pet health risks, diminished aquatic life, and economic impacts on tourism and recreational activities in addition to value drop of lakefront properties (Li et al., 2021).

3.2.2 In-Situ Chl-*a* Concentration Data

In-situ matching data on Chl-*a* concentration were extracted from the three databases identified in Table 3.2. These databases have varying data availability periods and Chl-*a* concentration measurement methods (Table 3.2). The datasets were filtered based on two criteria: (i) a maximum time window of ± 4 days between the in-situ measurements and satellite image acquisition dates, and (ii) a sampling water depth ranging from near-surface to a maximum of 1 m. A small portion of the datasets contained *pheophytin*-corrected records; consequently, the matchup analysis was conducted based on uncorrected Chl-a concentrations. In cases where uncorrected Chl-*a* concentration measurements were unavailable, *pheophytin*-corrected Chl-*a* data were utilized. This process resulted in 600 matchup data points for WLO and HH, covering the period from 2000 to 2022.

Of the total matchup data, 22% were collected during autumn/winter (September to February) and 78% in spring/summer (March to August). We used Carlson's TSI, which considers concentrations higher than 7.3 µg/L as characteristic of eutrophic to hypereutrophic waters and lower values as characteristic of oligotrophic to mesotrophic waters (Carlson, 1977). With this definition, 51% of the matchups fell in the oligotrophic/mesotrophic category, and the other 49% in the eutrophic/hypereutrophic category.

The mean Chl-*a* concentration across all matchups was 10.5 µg/L, with a standard deviation of 11.3 µg/L. See Table A1.2 in the supplementary materials (Appendix I) for further details on the in-situ Chl-*a* matchup data. A time-series plot of the in-situ data is also presented in Figure A1.1 of the supplementary materials colour-coded with their corresponding satellite matchups.

Table 3.2. Overview of in-situ data sources used in this study. ECCC = Environment and Climate Change Canada; MECP = Ministry of Environment, Conservation and Parks (Province of Ontario).

				Location		
				WLO + HH	**WLO**	**HH**
Source	**Organization**	**Published Data Availability**	**Chl-*a* Extraction Method**	**Fraction of Data (%)**	**Fraction within Study Site (%)**	
Hamilton Harbour Water Quality Data	ECCC	1987 – 2019	(National Laboratory for Environmental Testing, 2021)	68%	5%	98%
Great Lakes Nearshore - Water Chemistry	MECP	2000 – 2017	(MECP, 2015, 2016)	15%	43%	2%
Great Lakes Water Quality Monitoring and Surveillance Data	ECCC	2000 – Present	(Strickland & Parsons, 1972)	17%	52%	0%
				100%	100%	100%

In this chapter, in-situ Chl-*a* data were categorized into four categories (Figure 3.2): the 'All' category includes all the data, the location category separates data between WLO and HH, the seasonality category has 'Autumn/Winter (AW)' and 'Spring/Summer (SS)' subcategories, and the TSI category divides the data among the 'Oligotrophic/Mesotrophic (OM)' and 'Eutrophic/Hypereutrophic (EH)' subcategories. The categorization helps in evaluating the performance of each scheme and to better understand the factors that influence the accuracy of satellite-derived Chl-*a* concentration estimations. For instance, a chi-squared test from in-situ measurements revealed a significant association between location, and the TSI categories ($\chi^2 = 235$, $p =< 0.001$, $V = 0.63$), which reflects the typically more eutrophic conditions encountered in HH than WLO.

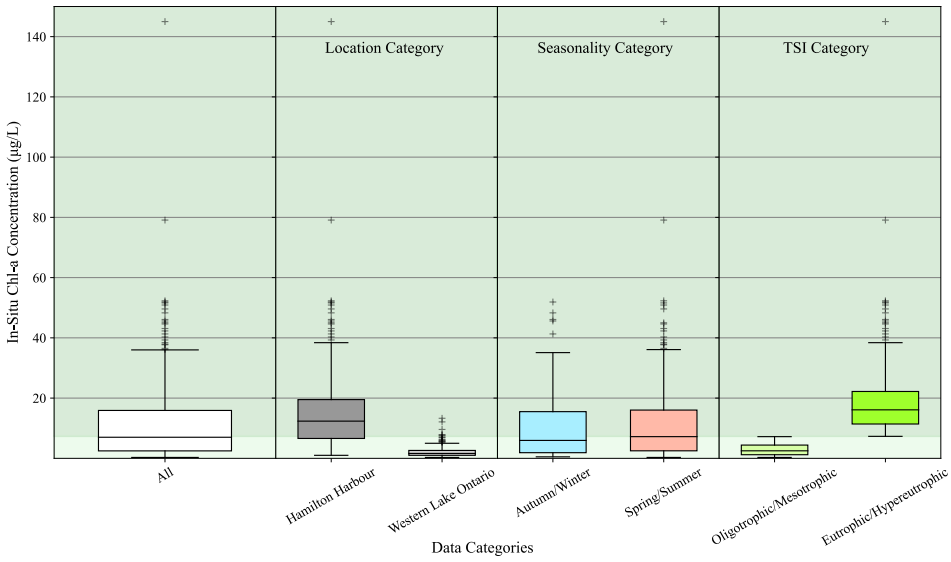

Figure 3.2. Boxplots of in-situ data, categorized by location, seasonality, and Carlson's Trophic State Index (TSI).

3.2.3 Remote Sensing Data

Landsat 5, 7, 8, and Sentinel-2 all capture multispectral images (Yan et al., 2022). Landsat 5, launched in 1984, provided data in seven spectral bands at a 30-meter resolution. Landsat 7, launched in 1999, introduced an additional 15-meter panchromatic band (Yin et al., 2023). Landsat 8, launched in 2013, enhanced the radiometric resolution and introduced nine spectral bands, including an ultra-blue and a cirrus band (O'Reilly & Werdell, 2019). The Sentinel-2 A/B constellation, launched in mid-2015, carries the multi-spectral instrument (MSI) with 13 spectral bands and variable spatial resolution from 10 to 60 meters (Warren et al., 2019). The differing spatial, spectral, radiometric and temporal resolutions of these satellites impact their capacity to estimate Chl-*a* concentration in water bodies (Beck et al., 2016; Lisboa et al., 2020). The literature suggests that the inclusion of an ultra-blue band at approximately 440 nm and the improved signal-to-noise ratio (SNR) of Landsat 8 and Sentinel-2 yields more accurate Chl-*a* concentration estimates compared to the previous generation satellites (Concha & Schott, 2016; Philipson et al., 2016). Table 3.3 summarizes the key features of each of the satellites mentioned above.

In total, 236 matchup scenes were selected after quality control for cloud, snow, shadow, and overlapping land pixel artefacts. The scenes were distributed as follows among the different satellites: Landsat 5: 79 scenes; Landsat 7: 89 scenes; Landsat 8: 49 scenes; and Sentinel-2: 19 scenes. Each scene contained at least one high-quality pixel coinciding with an in-situ measurement location. Most of the

downloaded RS scenes were acquired within ± 1 day of the in-situ sampling dates and at most within ± 4 days. Such close temporal matching is important given the rapidly changing algal patterns often observed in lakes (Adams et al., 2022).

Table 3.3. Overview of Landsat 5, 7, 8, and Sentinel-2.

		Landsat 5	**Landsat 7**	**Landsat 8 and 9**	**Sentinel-2 A/B**
Sensor		TM	ETM+	OLI and TIRS (OLI-2 and TIRS-2 for Landsat 9)	MSI
Operating Dates		1984 – 2013	1999 – Present	2013 – Present	2015 – Present
No. of Bands		7	8	11 (9 OLI, 2 TIRS)	13
Spatial Res. (m)		30 (15 for panchromatic band)	30 (15 for panchromatic band)	30 (15 for panchromatic band, 100 for TIRS)	10 (4 bands), 20 (6 bands), 60 (3 bands)
Temporal Res. (days)		16	16	8 (Landsat 8 and 9 combined)	~5 (Sentinel-2 A and B combined)
Radiometric Res. (bit)		8	8	12 (14 for Landsat 9)	12
Spectral Range (nm)		450 – 2,350	450 – 2,350	430 – 2,300 (OLI) 10,600 – 12,500 (TIRS)	443 – 2,190
Chl-*a* Retrieval Bands in nm (central wavelength)	λ_{UB}	-	-	430 – 450 (443)	433 – 453 (443)
	λ_{B}	450 – 520 (485)	450 – 520 (483)	450 – 515 (482)	457 – 523 (490)
	λ_{G}	520 – 600 (560)	520 – 600 (560)	525 – 600 (561)	542 – 578 (560)
	λ_{R}	630 – 690 (660)	630 – 690 (661)	630 – 680 (655)	650 – 680 (665)
	λ_{RE1}	-	-	-	697 – 713 (705)
	λ_{RE2}	-	-	-	732 – 748 (740)
	λ_{RE3}	-	-	-	773 – 793 (783)
	λ_{NIR1}	760 – 900 (830)	760 – 900 (835)	-	784 – 900 (842)
	λ_{NIR2}	-	-	845 – 885 (865)	935 – 955 (865)

Abbreviations in alphabetical order: G: Green, B: Blue, NIR: Near-Infrared, R: Red, RE: Red-Edge, TIRS: Thermal Infrared Sensor, UB: Ultra-Blue

3.2.4 Atmospheric Correction Processors

Atmospheric correction is a crucial step to accurately retrieve Chl-*a* concentration from satellite data because at-sensor radiance is affected by atmospheric gaseous molecules and aerosols, especially at shorter wavelengths (Tavares et al., 2021; Z. Wang et al., 2019), as well as by other factors such as air-water interface, wind, and sunglint (Pahlevan et al., 2020a; Pereira-Sandoval et al., 2019). While some atmospheric correction processors considered here can correct sunglint-related effects to a degree, an investigation of these effects is beyond the scope of our study.

The availability of atmospheric correction processors varies across the four satellites. For Landsat 5 and 7, both Level-1 and Level-2 (LEDAPS-corrected) products are publicly available, and the Level-1 imagery is also compatible with the ACOLITE, ATCOR, DOS1, FLAASH, and QUAC processors. In

the case of Landsat 8, Level-1 and Level-2 (LaSRC-corrected) are publicly available through the EarthExplorer data hub, and the satellite is compatible with all the above-mentioned processors plus C2RCC and iCOR. Finally, Sentinel-2, through the Copernicus data hub, freely offers Level-1 and Level-2 (Sen2Cor-corrected) products, also compatible with the ACOLITE, C2RCC, DOS1, iCOR, and Polymer processors. Although atmospherically corrected Sentinel-2 products (MSIL2A) are available for download, their global coverage only commenced in late 2018. Hence, we applied the Sen2Cor processors independently on Sentinel-2 Level-1 images using the ESA SNAP software. Table A1.3 in the supplementary materials summarizes the atmospheric correction methods utilized in this study.

3.2.5 Chl-*a* Retrieval Indexes

The Chl-*a* retrieval methods for satellite imagery broadly fall into four groups: empirical, semi-empirical, analytical, and semi-analytical (Pahlevan et al., 2020a; Tan et al., 2017). In this chapter, we focus on semi-empirical methods, which strike a balance between simplicity, reproducibility and speed as well as analytical accuracy (Dörnhöfer & Oppelt, 2016; Pirasteh et al., 2020). Semi-empirical algorithms for water quality assessment integrate water spectral theory with statistical techniques to establish relationships between the optically active constituents (OACs) in water and corresponding RS data (Buma & Lee, 2020). These approaches analyze the inherent optical properties (IOPs) of water constituents by selecting the most suitable method and band combinations for retrieving water quality parameters (Pirasteh et al., 2020; Yin et al., 2023). Semi-empirical models employ feature engineering, describing each pixel of an optical RS image through a numerical feature vector. This vector is then fed into a statistical or machine learning algorithm to develop a regression model for estimating water quality parameters, such as the Chl-*a* concentration (Gholizadeh et al., 2016; Ilteralp et al., 2022). Semi-empirical methods offer ease of implementation, reproducible results, and reasonable Chl-*a* concentration estimation accuracy without overfitting risk (Tan et al., 2017). The choice of semi-empirical methods is also dictated by the limited availability of in-situ matchup data for our study site that semi-analytical approaches depend on (Ogashawara et al., 2021).

Upon conducting an extensive literature review, 27 commonly used semi-empirical Chl-*a* indexes were selected. These indexes employ various band combinations to retrieve the Chl-*a* concentration from the matching vector of pixel values. The applicability of these indexes varies across different satellite platforms. Notably, Sentinel-2 and Landsat 8 demonstrate broader compatibility owing to their superior spectral resolution, particularly in ultra-blue, red-edge, and narrow near-infrared bands. The selected

indexes, together with the corresponding formulas (band math), names, references, and applicability, are found in Table 3.4. Next, these indexes were applied to atmospherically corrected images, selecting the most important feature and atmospheric correction processor, followed by regression analyses between the retrieved pixel values and in-situ Chl-*a* concentration.

Table 3.4. Chl-*a* retrieval indexes used in this study.

		Index Code	Band Math	Also Known As	Supported Sensors	Key References
2BDA	2-Band Ratios	I_1	$\frac{UB}{B}$	-	OLI, MSI	(Salem et al., 2017)
		I_2	$\frac{UB}{G}$	OC3E	OLI, MSI	(A. Gitelson et al., 1993; Kahru & Mitchell, 1998; Morel & Prieur, 1977; O'Reilly et al., 1998)
		I_3	$\frac{B}{G}$		TM, ETM+, OLI, MSI	
		I_4	$\frac{NIR1}{G}$	-	TM, ETM+, MSI	(A. A. Gitelson et al., 2002)
		I_5	$\frac{NIR2}{G}$		OLI, MSI	
		I_6	$\frac{B}{R}$	-	TM, ETM+, OLI, MSI	(Hewson et al., 2001; Zarco-Tejada & Ustin, 2001)
		I_7	$\frac{G}{R}$	-	TM, ETM+, OLI, MSI	(Guitelson et al., 1986; Lillesand et al., 1983; Yasuoka & Miyazaki, 1982)
		I_8	$\frac{RE1}{R}$	-	MSI	(Dekker & Peters, 1993; Dierberg & Carrlker, 1994; Gons, 1999; Mittenzwey et al., 1992; Moses et al., 2009)
		I_9	$\frac{RE2}{R}$	-	MSI	(A. A. Gitelson et al., 2009; A. A. Gitelson & Kondratyev, 1991; Koponen et al., 2007; Moses et al., 2009)
		I_{10}	$\frac{NIR1}{R}$	-	TM, ETM+, MSI	(Birth & McVey, 1968; A. A. Gitelson & Yacobi, 1995; Matthews, 2011; Moses et al., 2017)
		I_{11}	$\frac{NIR2}{R}$		OLI, MSI	
	Normalized Indexes	I_{12}	$\frac{G-R}{G+R}$	NDGRI	TM, ETM+, OLI, MSI	(Ahamed et al., 2011; Cheng et al., 2013; Feng & Hu, 2016; Tucker, 1979)
		I_{13}	$\frac{RE1-R}{RE1+R}$	NDCI	MSI	(S. Mishra & Mishra, 2012; S. P. Mishra & Jena, 2012)
		I_{14}	$\frac{RE2-R}{RE2+R}$	-	MSI	-
		I_{15}	$\frac{RE3-R}{RE3+R}$	-	MSI	-
		I_{16}	$\frac{NIR1-R}{NIR1+R}$	NDVI	TM, ETM+, MSI	(Rouse et al., 1973, 1974)
		I_{17}	$\frac{NIR2-R}{NIR2+R}$		OLI, MSI	
3BDA		I_{18}	$\frac{B-R}{G}$	BRG Index or KIVU	TM, ETM+, OLI, MSI	(Brivio et al., 2001; A. Gitelson et al., 1993; Kahru & Elmgren, 2014; Mayo et al., 1995; Mittenzwey et al., 1992)
		I_{19}	$RE1-\frac{R+RE2}{2}$	-	MSI	(Blaustein, 1992; Moses et al., 2009; Toming et al., 2016)
		I_{20}	$RE1-\frac{R+RE3}{2}$	-	MSI	(Kutser et al., 2016)
		I_{21}	$\left[\frac{1}{R}-\frac{1}{RE1}\right]\times RE2$	-	MSI	(Dall'Olmo et al., 2003; Dall'Olmo & Gitelson, 2006; Gordon et al., 1975; Moses et al., 2009)
		I_{22}	$\frac{\left[\frac{1}{R}-\frac{1}{RE1}\right]}{\left[\frac{1}{RE2}-\frac{1}{RE1}\right]}$	-	MSI	(A. A. Gitelson et al., 2008; Le et al., 2013; W. Yang et al., 2010)

FLH/RLH	I_{23}	$G - UB - (\frac{\lambda_G - \lambda_{UB}}{\lambda_R - \lambda_{UB}})(R - UB)$	CI	OLI, MSI	(C. Hu et al., 2012; Neville & Gower, 1977)
	I_{24}	$G - R - (\frac{\lambda_G - \lambda_R}{\lambda_B - \lambda_R})(B - R)$	-	TM, ETM+, OLI, MSI	(Neville & Gower, 1977)
	I_{25}	$RE1 - R - (\frac{\lambda_{RE1} - \lambda_R}{\lambda_{RE2} - \lambda_R})(RE2 - R)$	MCI or SLH	MSI	(A. Gitelson et al., 1994; J. Gower et al., 2005; J. F. R. R. Gower et al., 1999; Neville & Gower, 1977)
	I_{26}	$RE1 - R - (\frac{\lambda_{RE1} - \lambda_R}{\lambda_{NIR2} - \lambda_R})(NIR2 - R)$	MPH	MSI	(Matthews et al., 2012; Neville & Gower, 1977)
4BDA	I_{27}	$\frac{NIR2 - R}{B + G}$	SABI	OLI, MSI	(Alawadi, 2010; Candiani et al., 2005; Floricioiu et al., 2005)

Abbreviations in alphabetical order: 2BDA: 2-Band Algorithm, 3BDA: 3-Band Algorithm, 4BDA: 4-Band Algorithm, FLH: Fluorescence Line Height, MCI: Maximum Chlorophyll Index, MPH: Maximum Peak Height, NDCI: Normalized Difference Chlorophyll Index, NDGRI: Normalized Difference Green and Red Index, NDVI: Normalized Difference Vegetation Index, RLH: Reflectance Line Height, SLH: Scattering Line Height

3.2.6 Performance Metrics

Correlation was assessed using Pearson's r and Spearman's ρ coefficients. The former was calculated as:

$$r = \frac{\sum[(I-\bar{I})(\log(chla_{meas}) - \overline{\log(chla_{meas})})]}{\sqrt{\sum[(I-\bar{I})^2] \times \sum[(\log(chla_{meas}) - \overline{\log(chla_{meas})})^2]}} \quad (3.1)$$

where I is the spectral index (i.e., the independent variable or predictor) and $chla_{meas}$ is the measured in-situ Chl-a concentration (i.e., the dependent variable or response). Spearman's ρ coefficient was given by:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad (3.2)$$

where d is the difference between a pair of ranks and n is the number of data. For both correlation coefficients, the values range from -1 to 1. Statistically, a significant correlation exists when the absolute values of r and ρ fall within the range 0.5 to 1.

Regression analysis is often used in to assess the performance of semi-empirical RS-derived models against the water-truth parameters. Here, we employed the following metrics to assess the models' predictive capability (Barreneche et al., 2023; Cao et al., 2020; Pahlevan et al., 2020a):

$$RMSLE = \sqrt{\frac{1}{n}\sum\{[\log(chla_{meas}) - \log(chla_{mod})]^2\}} \quad (3.3)$$

$$RMSE\ (\mu g/L) = \sqrt{\frac{1}{n}\sum[(chla_{meas} - chla_{mod})^2]} \quad (3.4)$$

$$Bias = 10^{\{\frac{1}{n}\sum[\log(\frac{chla_{meas}}{chla_{mod}})]\}} \quad (3.5)$$

$$MAE = 10^{\left[\frac{1}{n}\sum\left|\log\left(\frac{chla_{meas}}{chla_{mod}}\right)\right|\right]} \quad (3.6)$$

$$MAPE\ (\%) = 100 \times \frac{1}{n}\sum\left[\frac{chla_{meas} - chla_{mod}}{chla_{meas}}\right] \quad (3.7)$$

$$MDAPE\ (\%) = 100 \times \tilde{r} \quad (3.8)$$

where $\tilde{r}$ is the median of $\frac{|chla_{mod}(i) - chla_{meas}(i)|}{chla_{meas}((i))}$ where $i = 1, \ldots, n$

$$\varepsilon\ (\%) = 100 \times (10^{Y} - 1) \quad (3.9)$$

where Y is the median of $\left|\log\left(\frac{chla_{meas}}{chla_{mod}}\right)\right|$

$$\beta\ (\%) = 100 \times (10^{Z} - 1) \times sing(Z) \quad (3.10)$$

where Z is the median of $\left[\log\left(\frac{chla_{meas}}{chla_{mod}}\right)\right]$

$$R^2 = 1 - \frac{\sum\{[\log(chla_{meas}) - \log(chla_{mod})]^2\}}{\sum\{[\log(chla_{meas}) - \overline{\log(chla_{meas})}]^2\}} \quad (3.11)$$

where the $\sum$ signs above denote summation over i =1, …, n. Generally, lower absolute values for the root mean squared logarithmic error (RMSLE), root mean squared error (RMSE), and mean absolute error (MAE) indicate better model performance, while higher values for R^2 imply better goodness of fit. Bias values near one indicate little bias, while values above and below one indicate average under- and over-prediction, respectively. Mean absolute percentage error (MAPE) and median absolute percentage error (MDAPE) measure the percentage error, with 0% implying perfect prediction. Similarly, the metrics ε and β report relative errors, with 100% indicating an accurate median prediction while values above (below) 100% indicate model overestimation (underestimation). Using these metrics together enabled a comprehensive model performance assessment. Note that, for some of the metrics a base-10 logarithmic transformation was used to enhance data normality and reliability (Pahlevan et al., 2020a). The overview of data and key steps of the methodology are summarized in Figure 3.3.

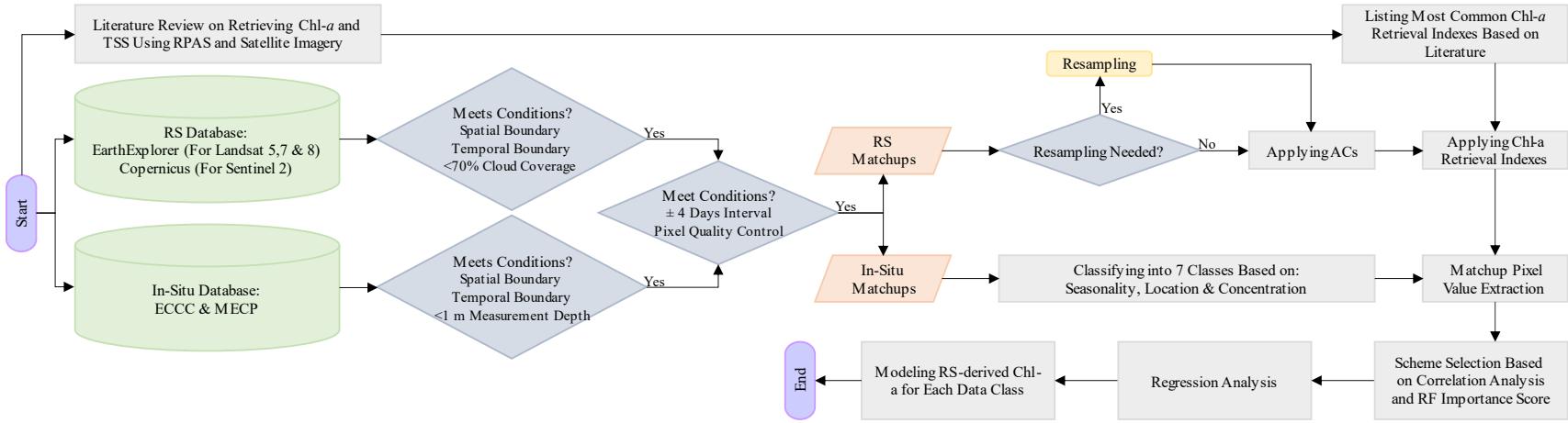

Figure 3.3. Flowchart of this study's methodology

3.2.7 Scheme Selection

In order to perform the regression analysis and develop an RS-derived Chl-*a* ($\mathrm{Chla_{RS}}$) model for a given scenario (i.e., a given satellite and data category combination), the primary (most significant) feature (i.e., the retrieval index) as well as atmospheric correction processor for the $\mathrm{Chla_{RS}}$ model were first identified. Random Forest (RF) scoring was used in the selection of the optimal scheme for each satellite and data category. Although in most cases, the scheme selection could be determined from the highest RF importance score, additional considerations arose when the top primary features showed nearly equal scores. Under these circumstances, correlation metrics were factored into the decision-making, as detailed in the next section.

3.3 Results

3.3.1 Feature Importance Scoring by RF

Feature importance analysis by RF ranks the significance of input variables in predicting the target output through a tree-based learning mechanism. In this chapter, RF was used to quantify the importance of the different combinations of retrieval indexes and atmospheric correction methods (i.e., the schemes) for the different data sources and categories (i.e., the different scenarios). The results, shown in Figure 3.4, exhibit notable differences in scheme importance across scenarios. While in some scenarios a scheme clearly dominated, indicating a distinct preference for that specific scheme, in other scenarios multiple schemes were of comparable importance. Among the atmospheric correction processors, QUAC and FLAASH tended to dominate for Landsat 5 and 8, Level-2 (LEDAPS) for Landsat 7, DOS1, C2RCC, and ACOLITE for Sentinel-2. Generally, FLAASH, DOS1, Level-2, and QUAC products performed best. Out of the 28 scenarios, Level-1 (uncorrected data) only emerged as the top product in 3 scenarios, hence, underscoring the critical importance of atmospheric correction prior to processing satellite data. Interestingly, iCOR and Polymer were never among the top processors, although they showed relatively high importance in several scenarios. In terms of the Chl-*a* indexes, I_3, I_4, and I_{18} were favoured by Landsat 5 and 7, I_1, I_3, I_6, and I_{24} by Landsat 8, and I_3 and I_{18} by Sentinel-2. Overall, I_3 emerged as the preferred index for Chl-*a* concentration prediction across all satellites and data categories.

Figure 3.4. Random Forest feature importance analysis with colour-coded atmospheric corrections processors. The x-axis denotes the retrieval index (feature), and the y-axis shows the importance score. For each scenario, the most significant scheme is marked with an asterisk.

3.3.2 Correlation and Regression Analysis

Correlation analysis was applied to examine relationships between in-situ Chl-*a* concentration and the corresponding spectral indexes derived from atmospherically corrected matchup pixels (Mortula et al., 2020; Sòria-Perpinyà et al., 2021). Figure 3.5 presents the resulting correlation coefficients for the retrieval indexes. A heatmap of only R^2 values from the correlation analysis is also provided in Figure A1.2 of the supplementary materials. These figures reveal that some indexes, such as I_{27}, exhibited an overall negative correlation, that is, an increase in the index value corresponds to a decrease in Chl-*a* concentration. Conversely, indexes such as for example I_{26}, exhibited a positive correlation for most scenarios, indicating a direct relationship between the index value and the Chl-*a* concentration. However, the majority of correlation metrics were in the range ±0.5, indicating statistically insignificant correlation. Furthermore, the results indicated that, regardless of the data category, both Sentinel-2 and Landsat 8 constantly had superior correlations compared to the other satellites. Of note is that while correlation analysis shows the strength and direction of the relationship, it does not provide insights into the models' prediction accuracy or biases.

Based on the results of the RF and correlation analyses, the optimal scheme (i.e., combination of atmospheric correction processor and Chl-*a* retrieval index) for each scenario was selected. Next, a linear regression was fitted between the scheme-derived and matchup Chl-*a* concentration to generate a predictive equation for Chl-*a*. The predicted versus matchup (measured) Chl-*a* concentration are plotted in Figure 3.6. Details on the performance metrics, as well as the regression equations are presented in Table A1.4 of the supplementary materials. The plots in Figure 3.6 are colour-coded by satellite (Landsat 5, 7, 8, and Sentinel-2, left to right) and identify the four in-situ data categories defined in section 3.2.2: All (no filtering), location (HH or WLO), seasonality (AW or SS), and TSI (OM or EH). The retrieval index and atmospheric correction processor, as well as the values for the matchup count, regression slope, RMSE, and R^2 are also shown on the plots.

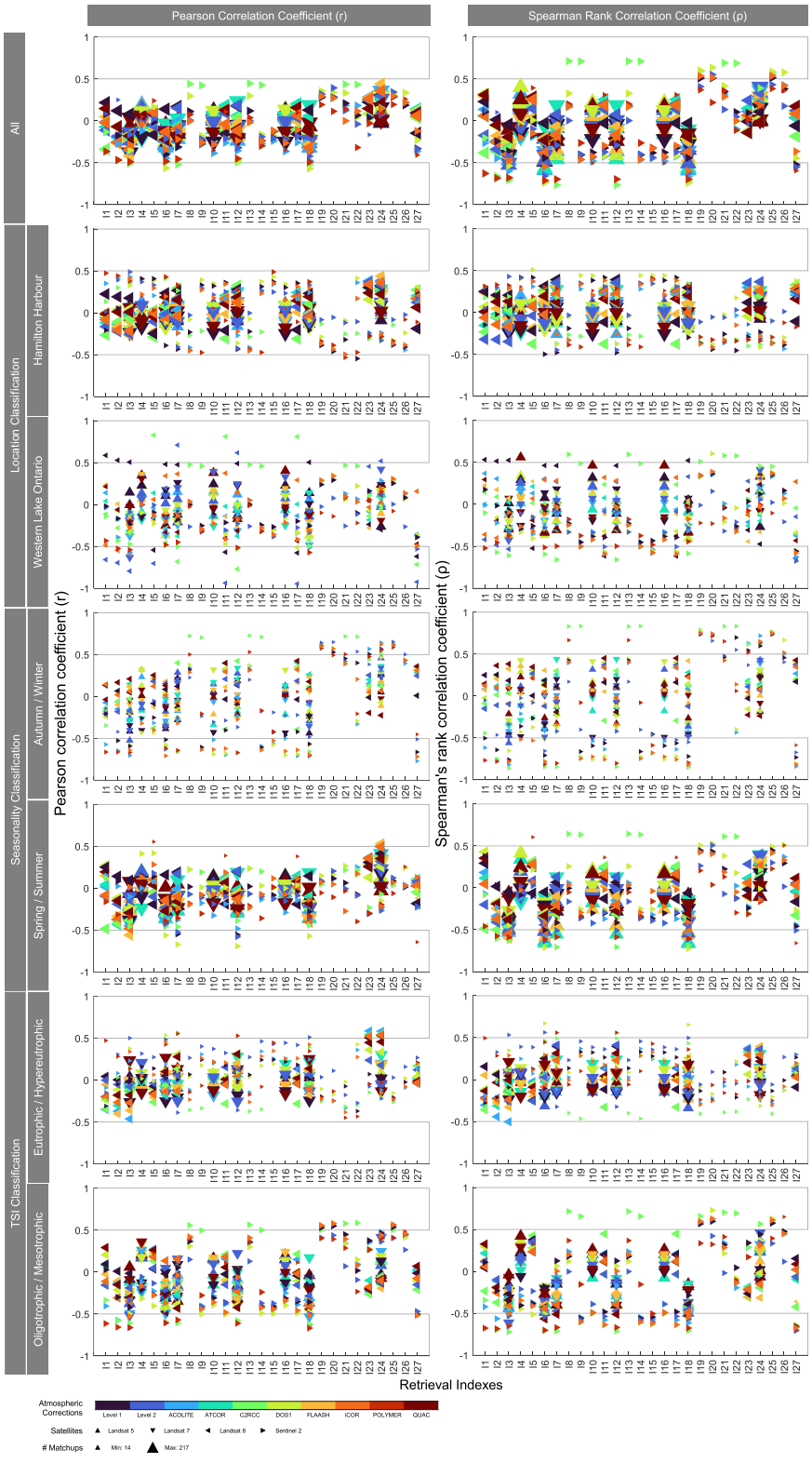

Figure 3.5. Evaluation of schemes across different scenarios based on correlation analysis. Marker colours denote different atmospheric corrections, shapes represent satellites, and sizes signify the number of matchups.

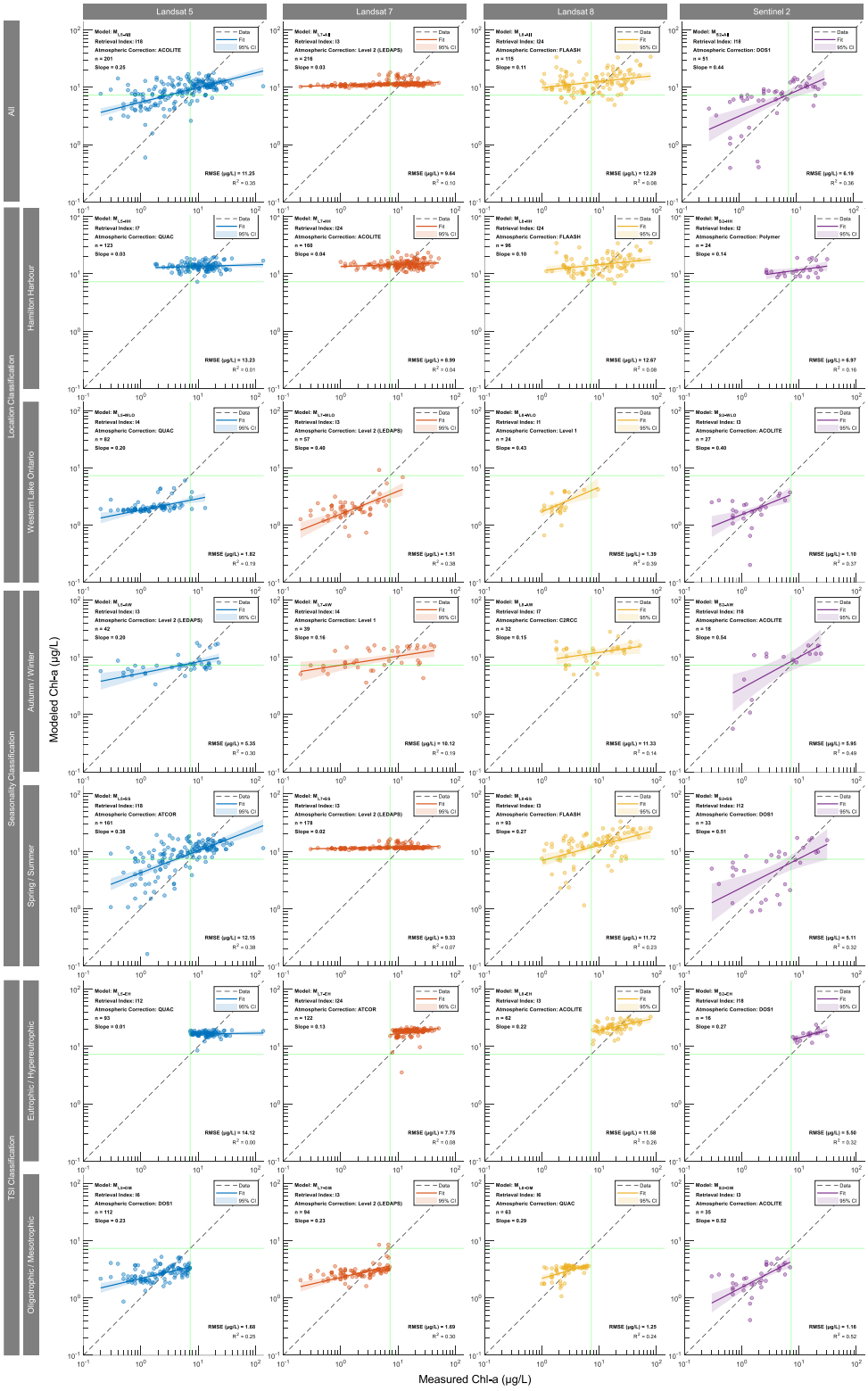

Figure 3.6. Plots comparing modeled vs. measured Chl-*a* concentration across satellites and data categories, demonstrating the regression models' performance

Landsat 5 and 7 data typically exhibited higher RMSE, from 1.5 (WLO) to 14.1 µg/L (EH). In contrast, Landsat 5's HH and EH categories had nearly zero slopes, indicating ineffective predictive power. The best performance for Landsat 5 occurred in the SS category, with a slope and R^2 of 0.38, indicating over- and under-prediction at low and high concentrations, respectively. The Landsat 7 models generally yielded low slopes, except for the WLO location subcategory model (slope: 0.40, R^2: 0.38, RMSE: 1.5 µg/L). Landsat 8 models had slopes between 0.11 and 0.43, and RMSE between 1.3 and 12.7 µg/L. The best Landsat 8 model for WLO data yielded a slope of 0.43, RMSE of 1.4 µg/L, and R^2 of 0.39. Sentinel-2 models generally outperformed those for the other satellites, with slopes from 0.14 to 0.54, RMSE from 1.2 to 7.0 µg/L, and R^2 from 0.16 to 0.52. We attribute this to Sentinel-2's superior spatial, spectral, and radiometric resolution. Most models for Sentinel-2 were atmospherically corrected using ACOLITE or DOS1. For the subcategory OM, Sentinel-2 exhibited the best performance (slope: 0.52, RMSE: 1.2 µg/L, R^2: 0.52) and, hence, provided the best option for the lower Chl-*a* concentration range due to its higher sensitivity (higher radiometric resolution) to variations in water colour.

3.4 Discussion

3.4.1 Performance of Satellites

Overall, compared to Landsat 5 and 7, Sentinel-2 and Landsat 8 demonstrate superior performance, with RMSLE values ranging from 0.18 ($M_{L8\text{-}WLO}$ and $M_{L8\text{-}OM}$) to 0.49 ($M_{L8\text{-}All}$) for Landsat 8, and 0.15 ($M_{S2\text{-}EH}$) to 0.38 ($M_{S2\text{-}All}$) for Sentinel-2. The average RMSLE for all seven scenarios is 0.34 for Landsat 8 and 0.27 for Sentinel-2, suggesting a slightly better performance for the latter. In contrast, Landsat 5 and 7 have RMSLE ranges of 0.21-0.41 and 0.18-0.52, respectively. The other performance metrics similarly imply the superior performance of Sentinel-2, followed by Landsat 8. For example, Sentinel-2, Landsat 8, 7, and 5 have R^2 ranges of 0.16-0.52, 0.08-0.39, 0.04-0.38, and 0.01-0.38, respectively. The comparable performance of Landsat 5 and 7 is consistent with the similar sensor configurations of both satellites. The weak performance of Landsat 5 and 7 is likely due to their lower SNR. It is also worth noting that the Sentinel-2 models were trained with less data than those for other satellites, mainly due to its more recent launch in mid-2015 and limited correspondence with in-situ measurements during the years of COVID-19 restrictions.

Among the four satellites, the Chl-*a* concentration predicted by the Landsat 7 models tend to exhibit the lowest slopes relative to the in-situ concentrations, averaging 0.14 across all seven data categories (Figure 3.6). The average slope is slightly better for Landsat 5 (0.19) and Landsat 8 (0.22), but distinctly superior

for Sentinel-2 (0.40). In principle, a perfect match between satellite-derived and in-situ measured Chl-*a* should yield a 1:1 slope. Given this, models such as $M_{L7\text{-}All}$, $M_{L7\text{-}HH}$, $M_{L7\text{-}SS}$, as well as $M_{L5\text{-}HH}$ and $M_{L5\text{-}SS}$, with slopes less than 0.04 can be considered entirely ineffective. However, even for the better performing Sentinel-2 models, the slopes are less than the theoretical 1:1 slope. This means that above and below a threshold value a given model under- and over-predicts the Chl-*a* concentration, respectively. For the HH plus EH categories this threshold is approximately 12.0 µg/L, while for the WLO plus OM categories it is about 1.2 µg/L. One factor contributing to the less than 1:1 slope could be the different spatial scales of the satellite versus in-situ Chl-*a* concentration. While the former average Chl-*a* concentration over entire pixels, the in-situ measurements sample the small-scale heterogeneity in phytoplankton distributions in the lake's surface waters. The range of the in-situ Chl-*a* (point) measurements can thus be expected to exceed that of the satellite-derived concentrations. On a model-predicted versus in-situ Chl-*a* concentration plot (Figure 3.6) this would translate in a slope of less than one.

3.4.2 Performance of Data Categories

Among the various categorization approaches, the 'All' category (no filtration) shows the poorest results, with RMSLE values as high as 0.49 and R^2 values as low as 0.08. Categorizing according to seasonality does not seem to be effective either, with the AW and SS categories exhibiting RMSLE values in the ranges 0.37-0.52 and 0.37-0.48, respectively. When in-situ data are categorized based on location, the WLO category performed moderately better, with RMSLE values ranging from 0.17 to 0.21, while the HH category showed slightly higher RMSLE values from 0.31 to 0.44.

The most promising results were obtained when in-situ data are categorized according to TSI. The EH and OM categories have relatively low average RMSLE values of 0.18 and 0.20, respectively. The superior performance for the EH category is not entirely surprising, given that the higher Chl-*a* concentration yield more intense spectral signatures (higher reflectance). This observation is supported by other performance metrics, such as MAE and MAPE.

3.4.3 Performance of Atmospheric Correction Processors

Among the final 28 top-performing models, only two were developed using Level-1 (i.e., without atmospheric correction) imagery, underscoring the vital importance of atmospheric correction as a preprocessing step. ACOLITE and LEDAPS feature prominently, with six and five top-performing models, respectively. Three out of the six ACOLITE-corrected models apply to Sentinel-2 imagery. Among the five

LEDAPS-corrected models, four are associated with Landsat 7 and one with Landsat 5. Regarding performance, the RMSLE for ACOLITE-corrected models ranges from 0.16 to 0.41, while for LEDAPS-corrected models, RMSLE ranges from 0.20 to 0.49.

Surprisingly, even though more sophisticated atmospheric correction processors specific to Sentinel-2 are available, the simpler DOS1 seems to be preferred for correcting Sentinel-2 data. Only for $M_{S2\text{-}HH}$ is the more advanced Polymer selected. Similarly, Landsat 8 models often perform better with simpler atmospheric correction processors, such as QUAC for $M_{L8\text{-}OM}$, or even with uncorrected Level-1 as in the case of $M_{L8\text{-}WLO}$. Nonetheless, more specialized atmospheric correction processors, like FLAASH (three times) or C2RCC, also rank among the top performers. By contrast, ATCOR is only preferred in two models, one for Landsat 5 and the other for Landsat 7, while none of the top-performing models use iCOR- or Sen2Cor-corrected data.

3.4.4 Performance of Retrieval Indexes

The I_3 index (blue-to-green ratio) scores highest, appearing nine times in the top-performing models. This corroborates literature reports of I_3's efficacy (Soriano-González et al., 2019). Of the nine instances, two were for Sentinel-2 and Landsat 8 each, four for Landsat 7, and one for Landsat 5. The second most frequent retrieval index among the 28 models is I_{18}, with five appearances, three of which are for Sentinel-2 with average RMSLE and R^2 values of 0.30 and 0.39, respectively. In third position, we find I_{24}, with two instances each for Landsat 8 and 7. The I_4, I_6, I_7, and I_{12} indexes each appeared twice, while I_1 appeared only once. Importantly, 18 of the 27 originally retained retrieval indexes from the literature are not included within the top-performing models. Notably, simple 2-band ratios appear to perform better than more complex 3-band ratios, while the sole 4-band index is not among the top performers. Overall, the most favoured combinations were the atmospheric correction processors ACOLITE paired with the I_3 retrieval index.

3.4.5 Performance of Individual Bands

Out of the final 28 top-performing models, the green band is included in 25 models, and the blue band in 21. Both bands feature prominently in retrieval indexes, for example I_3, I_{18}, and I_{24}. While their performance varies, they often yield promising results when paired with the red band, such as in $M_{S2\text{-}All}$, $M_{S2\text{-}AW}$, and $M_{S2\text{-}EH}$. The red band is utilized in 15 of the top-performing models and appears in various retrieval indexes, including I_6, I_7, and I_{12}. Despite the wide range of performance outcomes, the red band proves particularly

effective in some models, for example $M_{S2\text{-}SS}$. The near-infrared band, which is only used in I_4, appears in two models: one for Landsat 5 the other for Landsat 7. However, both models show unsatisfactory performance (R^2 of 0.19). Another band that appears in just two models is the ultra-blue, featured in I_1 and I_2 for $M_{L8\text{-}WLO}$ and $M_{S2\text{-}HH}$, with a better performance for $M_{L8\text{-}WLO}$ (RMSLE of 0.18 and an R^2 of 0.39). Contrary to expectations, the red-edge bands of Sentinel-2 do not appear in any of the retrieval indexes used in top-performing models. The most sensitive bands for modeling low concentrations of Chl-*a* are red, green and blue (RGB). Only at higher Chl-*a* concentration does a reflective peak in the red-edge become more distinguishable. The low Chl-*a* concentration training data likely explains why the red-edge bands were not used in the Sentinel-2 models: the highest Chl-*a* concentration for Sentinel-2 matchups is 17 µg/L.

3.4.6 Uncertainties

Several potential sources of uncertainty may impact the results presented. One significant source is the influence of various OACs, such as suspended solids and coloured dissolved organic matter (CDOM), which can directly affect water-leaving radiance and, therefore, can cause misestimations of Chl-*a* concentration (Pirasteh et al., 2020). Spatial heterogeneity of in-situ Chl-*a* concentration could be another source of uncertainty because small-scale variability of the Chl-*a* concentration across small spatial scales may not be captured at the spatial resolution of the satellite images (Le et al., 2013). Similarly, temporal mismatches between measurements and satellite data introduces uncertainties. Only 11% of the measurements were synchronous (i.e., on the same day) with the satellite overpass. Lastly, atmospheric errors are an unavoidable source of uncertainty in aquatic remote sensing studies. Despite significant improvements in atmospheric correction processors, none perfectly replicates the water-leaving radiance. Comparisons of calculated remote sensing reflectance with in-situ water-leaving reflectance show that even high-performing models can have median errors of up to 30% for the green and red bands, and up to 60% for the blue band (Pahlevan et al., 2021), with mean absolute differences of up to 60% (Warren et al., 2019).

3.5 Conclusions

This chapter assesses the influences of a variety of factors, including selection of satellites, in-situ data categories, atmospheric correction processors, and retrieval indexes, on the performance of RS-derived Chl-*a* models for the nearshore and offshore waters of the western basin of Lake Ontario, including Hamilton Harbour. Sentinel-2 is found to systematically outperform other satellites in Chl-*a* concentration retrieval. The results further emphasize the importance of categorizing in-situ Chl-*a* concentration data based on the

trophic state and tailoring algorithms to each category accordingly. Green and blue are found as the preferred bands for Chl-*a* concentration estimation. However, the effectiveness of spectral bands depends on the specific index, satellite, and atmospheric correction method used, which must be carefully evaluated in the development of Chl-*a* concentration retrieval models. Another key finding is that model complexity does not necessarily correlate with improved retrieval accuracy, implying that simpler models should be given appropriate consideration in RS water quality applications. This chapter adds to the literature on semi-empirical satellite-derived Chl-*a* retrieval approaches, that are emerging as essential tools in water quality monitoring that can help mitigate eutrophication issues in large freshwater lakes.

Chapter 4

Time-Series Analysis of Satellite-Derived Chl-*a* Concentration from 2013 to 2023 in Western Lake Ontario and Hamilton Harbour Using Landsat 8 and 9 Imagery

4.1 Introduction

The Laurentian Great Lakes, as North America's largest freshwater system, hold significant ecological, economic, and cultural value. They preserve biodiversity, support recreational and commercial activities, and are essential for drinking water supply (Mohamed et al., 2019). Lake Ontario, the easternmost of these lakes, lies between the Canadian province of Ontario and the US state of New York. Outflow from eutrophic Lake Erie via the Niagara River represent a major water and nutrient source (particularly of phosphorus, a key limiting factor) to the western basin of Lake Ontario (ECCC & U.S. EPA, 2022). The northern and western nearshore areas of the basin are further strongly influenced by the Greater Toronto Area (GTA) and the industrial city of Hamilton, respectively. These nearshore areas, including Hamilton Harbour, continue to face eutrophication challenges.

Eutrophication, due to nutrient over-enrichment, leads to algal blooms and reduced dissolved oxygen levels (Hiriart-Baer et al., 2009a; E. T. Howell & Benoit, 2021b; Sagan et al., 2020). High anthropogenic phosphorus loadings originate from wastewater and stormwater outflows (Chomicki et al., 2022), as well as non-point sources like urban runoff and agricultural fertilizer applications (Sagan et al., 2020). Harmful algal blooms (HABs) can disrupt a lake's ecological balance, causing biodiversity loss and socio-economic impacts, for instance by limiting activities and services like swimming, fishing, and water supply (Li et al., 2018). Studies have highlighted the occurrence of the nuisance benthic algae *Cladophora* (Dove & Chapra, 2015) from Burlington Beach to Humber Bay along Lake Ontario's north shore and harmful cyanobacterial blooms (cHABs) in Hamilton Harbour (ECCC & U.S. EPA, 2022). There is therefore an urgent need for effective management strategies to combat eutrophication and its adverse effects in Western Lake Ontario, including the three listed Areas of Concern (AOCs): Hamilton Harbour, Toronto waterfront, and the Niagara River.

Long-term water quality monitoring programs are essential for reporting on status and trends and identifying key factors that drive algal blooms (Tan et al., 2017). Effective monitoring requires data that reflects spatial and temporal variations across multiple sites (Mansaray et al., 2021). However, field observations often cover a limited number of sites over short durations due to the substantial costs involved in maintaining field personnel and equipment, as well as the subsequent laboratory analyses (Pokrzywinski

et al., 2022; Tan et al., 2017). Satellite-based monitoring offers a cost-effective alternative, delivering comprehensive observations with frequent overpasses, enabling the estimation of key water quality indicators such as Chlorophyll-*a* (Chl-*a*) concentration which serves as indicator of eutrophication (Lobo et al., 2021). This approach offers broader spatial and temporal coverage and is less labour-intensive compared to traditional ground-based measurements (Tuygun et al., 2023). When aligned with in-situ data, satellite-derived models yield time-stamped, geolocated data that can identify trends and irregularities beneficial for detecting (re)emerging productivity hotspots (Sagan et al., 2020).

The optical response of phytoplankton biomass to solar radiation energy is significantly correlated with the concentration of Chl-*a* (Cao et al., 2020; Tan et al., 2017; Xiao et al., 2023). This photosynthetic pigment indicates the presence of phytoplankton, including cyanobacteria (Chegoonian et al., 2023), and can serve as an indirect indicator of nutrient status (Leggesse et al., 2023). Water colour remote sensing is based on the fact that changes in the concentration of photosensitive parameters, like Chl-*a*, lead to alterations in the optical properties of a water body, consequently changing its reflectivity (Yuan et al., 2022). Because Chl-*a* pigments exhibit distinct spectral signatures, their concentration can be estimated using remote sensing data (Souza et al., 2021). By integrating in-situ near-surface concentrations of Chl-*a* with reflectance data from coincident satellite overpasses, a regression-based models can be produced (Khan et al., 2021). Such models can then be extrapolated to yield Chl-*a* concentration for images without matchups, proving advantageous for time-series analysis of algal blooms, especially when field data gaps exist.

The availability of various satellites, each with different spatial, spectral, radiometric, and temporal resolutions, enables time-series analysis of Chl-*a* concentration. Key among these are medium-spatial-resolution (< 1000 m) satellites (sensors in parenthesis) such as Sentinel-3 (OLCI), Aqua/Terra (MODIS), EnviSAT (MERIS), and Suomi NPP/ NOAA-20 (VIIRS), OrbView-2 (SeaWiFS), as well as high-spatial-resolution satellites (< 100 m) like Landsat 1 to 9 (MSS, TM, ETM+, OLI, OLI-2), and Sentinel-2 A/B (MSI) (Tavora et al., 2023). High-spatial-resolution sensors, such as OLI (30 m) typically have longer revisited periods (16 days, halved when combined with OLI-2) compared to medium-spatial-resolution counterparts like MODIS and OLCI (250-500 m), which offer near-daily coverage but with a larger pixel size. Landsat and Sentinel-2 constellations have been more frequently used for small inland lakes, while MODIS is popular for coastal and oceanic waters due to its moderate spatial resolution and optimal revisit time (Khan et al., 2021; Tavora et al., 2023). The Landsat series, initiated with the launch of Landsat 1 in 1972, has progressed to Landsat 9, launched in 2021, which features capabilities nearly identical to its predecessor, Landsat 8, launched in 2013. To maintain consistency in the satellite-derived Chl-*a* concentration products, we exclusively utilized Landsat 8 and 9 data, resulting in a combined revisit time of 8 days, 30-m spatial resolution and five optical bands ranging from ultra-blue to near-infrared. Due to

the low SNR of the TM and ETM+ sensors, which makes them unfit for retrieving water quality products (Tavora et al., 2023), as well as the appearance of Landsat 7's ETM+ scan line error in 2003, we did not include Landsat 5 or 7 imagery in this study. Additionally, despite offering a finer 10-meter spatial resolution since mid-2015, we decided against employing Sentinel-2 (MSI) data because of their different spectral characteristics, and the computational demands of processing multiple tiles for mosaicking. Interestingly, a recent similar study by Niroumand-Jadidi et al. (2022) reported that Landsat 9 images exhibited less grainy noise and relatively higher accuracies in matchup analysis compared to Sentinel-2.

Numerous recent studies have estimated Chl-*a* concentration of inland waters using Landsat 8 and the recently launched Landsat 9 (Niroumand-Jadidi et al., 2022) imagery. Some of these studies emphasize developing and testing retrieval algorithms and assessing their performance (Chegoonian et al., 2023; Villota-González et al., 2023) while others analyze spatial-temporal trends over short periods in one (Poddar et al., 2019; D. Wang et al., 2022) or several (Boucher et al., 2018; Markogianni et al., 2018) case studies. However, only a few of these studies have a temporal extent enabling meaningful trend analyses. Existing studies with time-series analyses of Chl-*a* using similar sensors and with at least two years of data are briefly compared with our study in Table 4.1.

The objective of this chapter is to generate Chl-*a* concentration maps for Western Lake Ontario and Hamilton Harbour from 2013 to 2023, using Landsat 8 and 9 imagery, with a focus on the contrasting conditions in nearshore and offshore waters. The study, validated through in-situ data matchups, examines the spatial and temporal dynamics of algal blooms using the methodology outlined by Binding et al. (2021) to calculate algal bloom intensity, extent, and severity. We further explore trends and hotspots in Chl-*a* concentration based on these indicators and analyze trends in annually and monthly averaged estimated Chl-*a* concentration maps derived from the satellite data.

4.2 Materials and Methods

The overall methodology is summarized in Figure 4.1. It encompasses the in-situ and remote sensing data collection, atmospheric correction via ACOLITE, implementation of the Chl-*a* retrieval model, evaluation through performance metrics, and estimation and analysis of algal bloom indicators.

4.2.1 Study Area

Lake Ontario, situated as the easternmost and terminal lake in the Laurentian Great Lakes (LGL) chain, is the smallest LGL in terms of surface area (19,000 km²) but still ranks as the 13th largest lake globally. The lake has a 6-to-8-year water residence time. It receives water primarily from upstream Lake Erie via the Niagara River, supplemented by inflow from multiple rivers and canals. The lake discharges at its eastern end into the St. Lawrence River. Lake Ontario serves as a drinking water source for nine million people and accommodates 36 water treatment plants along its shores (ECCC & U.S. EPA, 2022). It receives nutrients from multiple sources, including agricultural and urban runoff, municipal wastewater treatment and industrial effluents, and stormwater outflows (E. T. Howell & Benoit, 2021a, 2021b).

Here, we focus on the western section of Lake Ontario (Figure 4.2), which includes three Areas of Concern (AOCs): the Toronto waterfront, the Niagara River outflow area, and Hamilton Harbour. The latter is a shallow, semi-enclosed embayment connected to Lake Ontario via a shipping channel. It spans 21.5 km² and reaches depths up to 24 m (Munawar & Fitzpatrick, 2018). Compared to other nearshore zones of Lake Ontario, Hamilton Harbour is characterized by calmer waters and warmer summer temperatures (Munawar et al., 2017), which favour cHABs and the development, and persistence, of hypoxia and anoxia in the hypolimnion (Zastepa & Chemali, 2021). The algal blooms in Hamilton Harbour and the other AOCs cause significant socio-economic (including tourism, real estate, and water quality) and ecological (including food web resilience, and habitats) costs (Carmichael & Boyer, 2016).

Our study area extends from the western end of Lake Ontario eastwards to Oshawa, Canada, and Olcott, USA. Some of the nearshore regions are considered mesotrophic (E. T. Howell & Benoit, 2021a, 2021b), the offshore waters are mostly oligotrophic (Dove & Chapra, 2015) and hypereutrophic conditions are observed for Hamilton Harbour (Hiriart-Baer et al., 2009a). Harmful algal blooms (HABs) in Hamilton Harbour are dominated by cyanobacteria, while the northern Mississauga-Oshawa waterfront, especially Toronto Harbour, are dominated by *Cladophora* growing on the shallow lakebed to depths where light can penetrate (Anderson et al., 2019; E. T. Howell & Benoit, 2021b; Munawar & Fitzpatrick, 2018). *Cladophora* provides habitat for aquatic life but its excessive growth can have negative economic impacts (e.g., on lakefront real estate prices), impair recreational activities (e.g., swimming, boating and fishing) and clog water intakes (E. T. Howell & Benoit, 2021b; U.S. EPA & Government of Canada, 2022).

The study area's bathymetry and surrounding land use and land cover are shown in Figure 4.2. The figure also highlights the nearshore waters, defined as littoral zones with a depth ≤ 15 meters (Binding et al., 2015), which represent the typical *Cladophora* habitat where growth is both phosphorus- and light-limited (Jabbari et al., 2023).

Table 4.1. Comparative summary of recently-published time-series studies on Chl-*a* retrieval using moderate-to-high-spatial-resolution multispectral imagery (top-performing methods and features are identified in bold).

Reference	Study Area	Temporal Coverage	Chl-*a* Matchups			Satellite Matchups			Retrieval Algorithm		
			Time Window	Total Count*	Range (µg/L)	Sensors Used	Atmospheric Correction	Radiometric Products	Method	Selected Features*	Performance Metrics*
(Tan et al., 2017)	Erhai Lake (China)	1987–2016	±8 d	73	3.6 – 46.7	TM, ETM+	LEDAPS	R_{RS}	SE (Linear, **MLR**)	Various bands and band combinations ($(\frac{\mathbf{1}}{\boldsymbol{\rho(485)}} - \frac{\mathbf{1}}{\boldsymbol{\rho(560)}}) \times \boldsymbol{\rho(830)}$)	RMSE = 0.10, PRMSE = 4.1%
(Guo et al., 2022)	Grand Lake and Hudson Reservoir located (USA)	2013–2017 2015–2020	±5 d	1709	0 – 60	OLI MSI	LaSRC Sen2Cor	ρ_S	ML (MR, SVR, **RFR**)	Various bands and band combinations ($\boldsymbol{\rho(10895)}$, $\boldsymbol{\rho(655)}$, $\boldsymbol{\rho(560)}$)	RMSE = 0.36
(Lobo et al., 2021)	Multiple Water Bodies (Brazil)	2015–2020	±2 d	136	0.6 – 435.7	MSI	SIAC	ρ_S	SE (Polynomial and **Power-Law)**	$\frac{\rho(705) - \rho(665)}{\rho(705) + \rho(665)}$	MAPE = 90%
(Maciel et al., 2023)	Río de la Plata estuary (South America)	2018–2020	–	43	0.4 – 177.6	MSI	ACOLITE	ρ_w	SE (Quadratic)	Various band combination ($\frac{\boldsymbol{\rho(705)-\rho(665)}}{\boldsymbol{\rho(705)+\rho(665)}}$)	RMSE = 4.9
(Chegoonian et al., 2023)	Buffalo Pound Lake (Canada)	2014–2020	±3 d	193	1 – 125	OLI, MSI	ACOLITE, **iCOR**	$\boldsymbol{R_{RS}}$, ρ_w, ρ_{TOA}	SE (Linear) and ML (MDN, **SVR**)	$\frac{\boldsymbol{\rho(705)}}{\boldsymbol{\rho(665)}}$ etc. All VIS-NIR bands	RMSE = 13.96, MAPE = 31.5
(Mansaray et al., 2021)	Multiple Water Bodies (USA)	2017–2020	±2 d	100 129 207	0.6 – 540	OLI MSI - (PlanetScope)	LaSRC Sen2Cor 6S	ρ_S	SE (Multivariate Quadratic Multivariate Linear Multivariate Linear)	Various bands and band combinations	RMSE = 2.04 RMSE = 1.19 RMSE = 4.41
(Hafeez et al., 2019)	Multiple Water Bodies (China)	1999–2015	Same day	120	0.3 – 28	TM, ETM+, OLI	6S	ρ_S	SE (Linear) and ML (SVR, RFR, **ANN**, CB)	Various bands and $\frac{\rho(655)}{\rho(485)^2}$, $\frac{\rho(865)}{\rho(485)^2}$, $\frac{\boldsymbol{\rho(655)}}{\boldsymbol{\rho(485)^2}}$, $[\rho(485) - \rho(865)]$	RMSE = 1.94, MAE = 1.48 RMSE = 1.4, MAE = 1.13
(Cao, Ma, Liu, et al., 2022)	Multiple Water Bodies (China)	2013–2022	±3 h and ±3 d	74 105	0.1 – 180	OLI MSI	– (in SeaDAS), **ACOLITE**, C2RCC, Polymer	R_{RS}	SE and ML (**MDN**, RF, SVR)	All VIS-NIR bands	MAPE = 31.4% MAPE = 38.0%
(Yin et al., 2023)	Lake Taihu (China)	1984–2021	±1 d	31 44	1 – 65 1 – 83	TM OLI	LEDAPS LaSRC	R_{RS}	SE (Exponential)	$\frac{\rho(655)}{\rho(865)}$ (Various band combinations tested)	RMSE = 9.57, AURE = 36.7% RMSE = 11.5, AURE = 38.1%
(Arias-Rodriguez et al., 2021)	Multiple Water Bodies (Mexico)	2013–2019	±3 d	40 32 20	0 – 32 0 – 37 0 – 30	OLI OLCI MSI	LaSRC C2RCC C2RCC	ρ_S R_{RS} R_{RS}	SE (Linear) and ML (SVM, **ELM**)	All VIS-NIR bands	RMSE = 19.99, MAE = 13.81 RMSE = 21.27, MAE = 8.86 RMSE = 8.47, MAE = 6.23

(Cao, Ma, Melack, et al., 2022)	Lake Taihu (China)	1984–2019	±3 h	43 95 96	1.4 – 99.1	TM ETM+ OLI	ACOLITE	R_{RC}	SE and ML (SVR, RFR, **XGB**)	Various bands and band combinations	RMSE < 16, MAPE < 35%
(Sagan et al., 2020)	Multiple Water Bodies (USA)	2013–2018	–	96	0.6 – 79.5	OLI, MSI	C2RCC	ρ_S	SE (PLSR) and ML (SVR, DNN) and Inversion Models (WASI, **NN Inversion**)	Various band combinations (four band rations and four shape algorithms)	RMSE = 2.02, MAPE = 11.66%
(Leggesse et al., 2023)	Lake Tana (Ethiopia)	2016–2022	–	931	0 – 191.6	OLI	LaSRC	ρ_S	ML (ANN, SVR, RFR, **XGB**, AB, GB)	Various bands and band combinations (different depending on the model but range from 12 for AB and XGB up to 87 for ANN)	RMSE = 9.79, MARE = 0.082
(Cao et al., 2020)	Multiple Water Bodies (China)	2013–2018	±6 h	225	0.04 – 258.7	OLI	–	R_{RC}	SE and ML (RF, **XGB**)	Various bands and band combinations ($\frac{\rho(443)}{\rho(561)}$, $\frac{\rho(655)}{\rho(561)}$, $\frac{\rho(865)}{\rho(655)}$, $\rho(865)-\rho(655)+[\frac{865-655}{1609-655}]\times[\rho(655)-\rho(1609)]$)	RMSE = 7.1, MAPE = 24%
(Yao et al., 2021)	Beibu Gulf (China)	2013–2020	Same day	110	2 – 15	OLI	FLAASH	ρ_S	SE and ML (ANN, SVM, **GBDT**)	Various bands and band combinations	RMSE = 1.63, MAPE = 19.4%
(F. Zhang et al., 2021)	Guanting Reservoir (China)	1985–2019	–	102	5.0 – 38.7	TM, ETM+, OLI (inter-sensor normalization)	ACOLITE	R_{RS}	SE (Linear)	$\frac{\rho(655)}{\rho(482)}$, $\frac{\rho(865)}{\rho(655)}$, $\frac{\rho(655)}{\rho(562)}$, $\frac{\rho(482)-\rho(655)}{\rho(562)}$, $\frac{\frac{1}{\rho(655)}-\frac{1}{\rho(482)}}{\rho(865)}$, $\frac{\boldsymbol{\rho(865)-\rho(482)}}{\boldsymbol{\rho(865)+\rho(482)}}$	RMSE = 5.67, MRE = 25.7%
(He et al., 2021)	Yangtze River (China)	2014–2020	±5 d	55	0.5 – 7.5	OLI	LaSRC	ρ_S	SE (Quadratic)	Various band combinations ($\frac{\boldsymbol{\rho(482)}}{\sqrt{\boldsymbol{\rho(443)^2+\rho(482)^2+\rho(865)^2}}}$)	RMSE = 0.47, MAPE = 25.8%
(X. Yang et al., 2020)	Donghu Lake (China)	1987–2018	±1 d	65	10 – 160	TM, ETM+, OLI	FLAASH	ρ_S	SE (Linear, Power, Exponential, Logarithmic and **Cubic**)	Various bands and band combinations ($\frac{\boldsymbol{\rho(562)-\rho(655)}}{\boldsymbol{\rho(562)+\rho(655)}}$)	RMSE = 11.19†, R^2 = 0.86†
(Yuan et al., 2022)	Pearl River Estuary (China)	2014–2021	–	25	0 – 400	OLI	FLAASH	ρ_S	SE (Linear, **Quadratic**, Logarithmic, Exponential)	Various bands and band combinations ($\frac{\boldsymbol{\rho(865)}}{\boldsymbol{\rho(655)}}$, $\frac{\rho(865)-\rho(655)}{\rho(865)+\rho(655)}$)	RMSE = 1.21, MAPE = 15.3%
Current Study	Western Lake Ontario (Canada)	2013–2023	±2 d	51	0.5 – 46.0	OLI and OLI-2	ACOLITE	ρ_w	SE (Linear, Quadratic, Cubic, **Exponential**) and ML (MDN)	Various band combinations ($\frac{\boldsymbol{\rho(482)}}{\boldsymbol{\rho(655)}}$)	RMSE = 8.49, MAPE = 78.8%

Abbreviations in alphabetic order: 6S = Second Simulation of a Satellite Signal in the Solar Spectrum, AB = AdaBoost Regression, ACOLITE = Atmospheric Correction for OLI 'lite', ANN = Artificial Neural Networks, AURE = Average Unbiased Relative Error, C2RCC = Case 2 Regional Coast Colour, CB = Cubist Regression, ConvLSTM = Convolutional Long Short-Term Memory, DL = Deep Learning, DNN = Deep Neural Networks, ELM = Extreme Learning Machine, ETM+ = Enhanced Thematic Mapper Plus, FLAASH = Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes, GB = Gradient Boosting, GBDT = Gradient Boosting Decision Tree, iCOR = Improved Coupled Ocean-Atmosphere Radiative Transfer, LaSRC = Landsat Surface Reflectance Code, LEDAPS = Landsat Ecosystem Disturbance Adaptive Processing System, MAE = Mean Absolute Error, MAPE = Mean Absolute Percentage Error, MARE = Mean Absolute Relative Error, MDN = Mixture Density Networks, ML = Machine Learning, MLR = Multivariate Linear Regression, MSI = Multispectral Instrument, MR = Multiple Regression, NN = Neural Networks, OLCI = Ocean and Land Colour Instrument, OLI = Operational Land Imager, PLSR = Partial Least Squares Regression, RFR = Random Forest Regression, RMSE = Root Mean Square Error, SE = Semi-Empirical, SIAC = Surface Information in the Atmospheric Correction, SVR = Support Vector Regression, TM = Thematic Mapper, VIS-NIR = Visible and Near-Infrared, WASI = Water Colour Simulator, XGB = Extreme Gradient Boosting

* For the top-performing models. The performance metrics are between observed vs predicted Chl-*a* (otherwise noted). The unit of RMSE is µg/L.

† Between observed Chl-*a* and the selected feature(s)

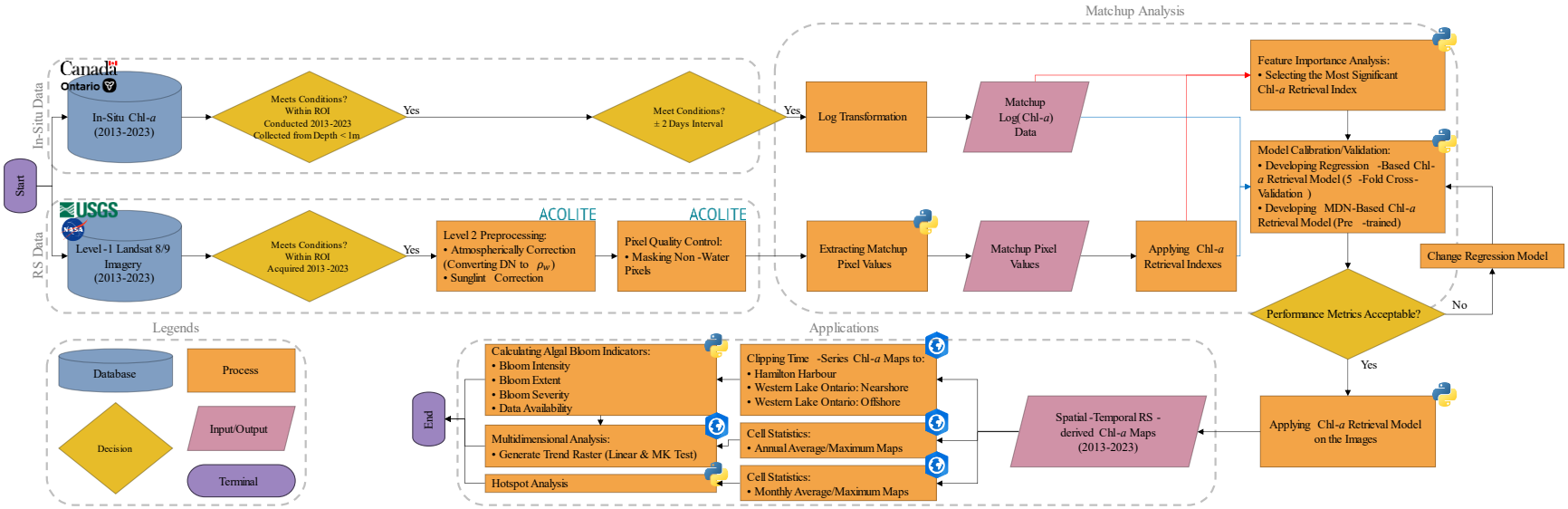

Figure 4.1. Flowchart of methodology

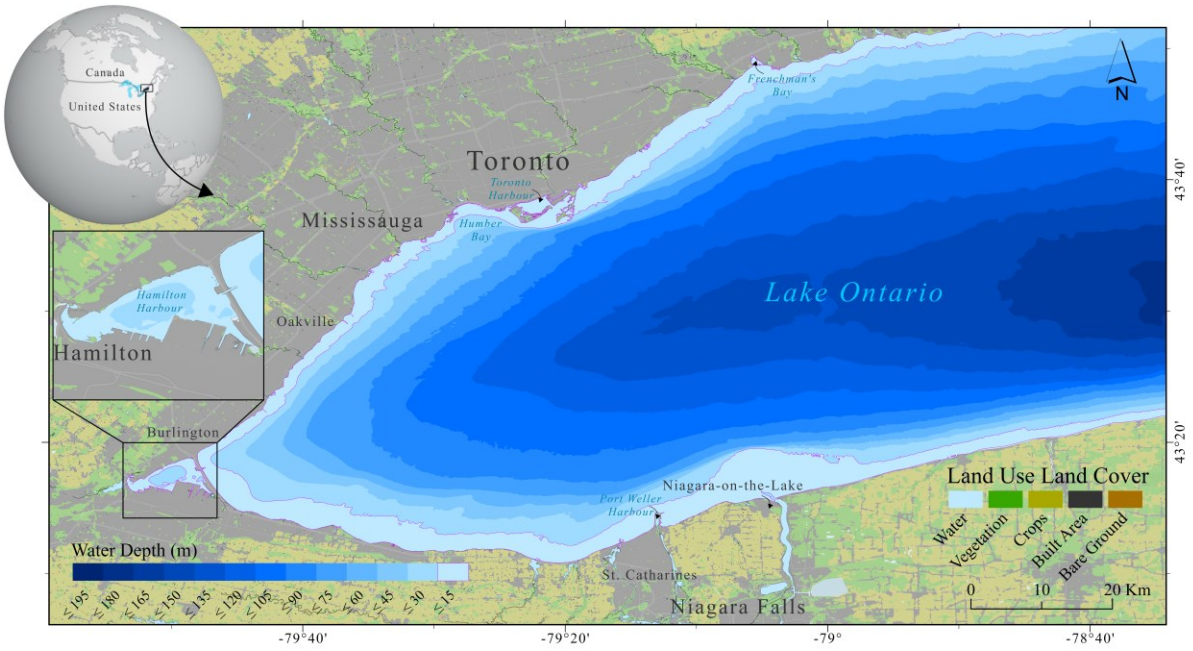

Figure 4.2. Map of the study area

4.2.2 In-Situ Chl-*a* Concentration Data

Calibrating and validating Chl-*a* retrieval models, whether using semi-empirical or machine learning (ML) methods, requires in-situ matchups for ground-truthing or in this case water-truthing (a commonly used misnomer as field data carry their own errors). We examined various water quality databases from binational academic and government sources to gather in-situ Chl-*a* concentration data from 2013 to 2023. The data underwent screening, selecting measurements from near-surface depths (≤ 1 m) and taken within ±2 days of satellite overpass, with available water pixel coverage in a 3 × 3 window around the measurement locations. Preference was given to *pheophytin*-corrected records, using uncorrected data only when corrected versions were unavailable. This yielded 51 matchup points, summarized in Table 4.2, and illustrated in Figure 4.3.

Table 4.2. Description of the in-situ data used for the matchup analysis

Source	Publisher	Chl-*a* Extraction Method	Pigment Correction	Spatial Coverage	Matchups	
					Counts	Frequency
Hamilton Harbour Water Quality Data (Access Link)	ECCC	(National Laboratory for Environmental Testing, 2021)	Pheophytin-Corrected	HH	17	33%
Water Chemistry (Great Lakes Nearshore Areas) (Access Link)	MECP	(MECP, 2023)	Pheophytin-Corrected	WLO Offshore and HH	11	22%
Great Lakes Water Quality Monitoring and Surveillance Data (Access Link)	ECCC	(Strickland & Parsons, 1972)	Uncorrected	WLO Nearshore and Offshore	10	20%
Hamilton Harbour Area of Concern (AOC) Monitoring (Access Link)	MECP	(MECP, 2023)	Uncorrected	HH	8	16%
A national harmonized dataset of discrete chlorophyll from lakes and streams (2005-2022) (Access Link)	USGS	(Arar & Collins, 1997)	Pheophytin-Corrected	WLO Nearshore	5	10%
				Total	51	100%

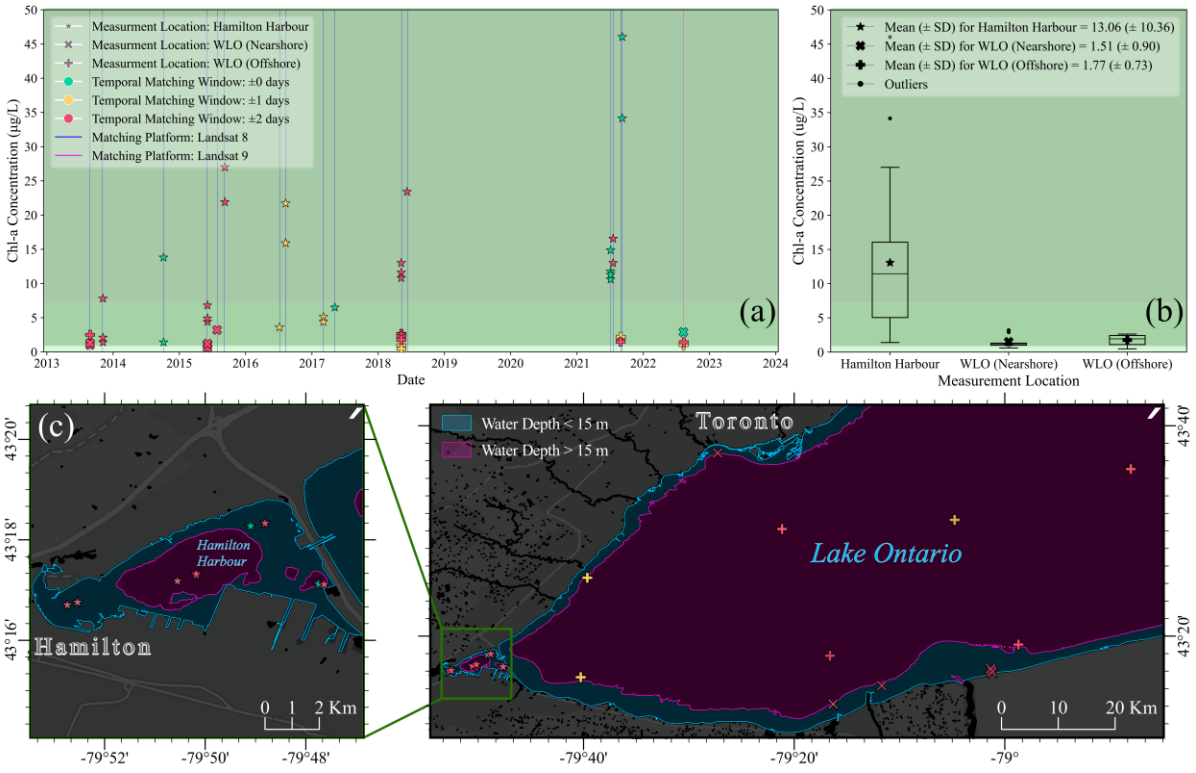

Figure 4.3. Temporal and spatial distribution of in-situ Chl-*a* concentration matchups. (a) and (c) represent the temporal and spatial distribution of the matchups, respectively while (b) visualizes data ranges for each region in box plots.

4.2.3 Remote Sensing Data

Landsat 8 and 9, launched in 2013 and 2021, respectively, by NASA and USGS, orbit in a Sun-synchronous, near-polar circular path at an altitude of 705 km, covering a scene size of 185 km cross-track by 180 km along-track. Landsat 8 is equipped with the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS), whereas Landsat 9 carries the upgraded OLI-2 and TIRS-2. Both OLI and OLI-2 are push-broom sensors, with significant advancements in signal-to-noise ratio (SNR) and radiometric resolution compared to the Enhanced Thematic Mapper (ETM+) sensor of Landsat 7 (Cao et al., 2020; Niroumand-Jadidi et al., 2022). Additionally, both include nine spectral bands with a 30 m spatial resolution, along with a 15-m panchromatic band. The central wavelengths of both sensors are almost identical, including bands at 443, 483, 561, 655, 865, 1609, 2201, 592, and 1373 nm and offer radiometric resolutions of 12 and 14 bits, capable of capturing 4096 and 16384 shades of a given wavelength, respectively (Chegoonian et al., 2023; Niroumand-Jadidi et al., 2022) allowing detailed detection of subtle variations in water-leaving radiance.

In this chapter, a total of 420 Level-1 Terrain Corrected Product (L1TP) Landsat collection 2 images (359 from Landsat 8 and 61 from Landsat 9) covering the period from March 2013 to September 2023 were downloaded from the USGS EarthExplorer data hub, encompassing all available images without a cloud cover filter pertaining to path 17 and 18, row 30, covering the western part of Lake Ontario. The L1TP scenes have already undergone radiometric and geometric (orthorectification) corrections by employing Ground Control Points (GCPs) and Digital Elevation Model (DEM) data, for topographic displacement, thereby providing premier quality Level-1 products apt for pixel-level time-series analysis. The next step was to mask non-water pixels and apply atmospheric correction – converting Level-1 digital numbers (DN) to water-leaving reflectance (ρ_w) – using the ACOLITE processor. Sections 1 and 2 in the supplementary materials (Appendix II) elaborate on both steps in more detail. After applying quality masks and generating Level-2 products, a total of 342 usable images remained for subsequent application of Chl-*a* retrieval model.

4.2.4 Atmospheric Correction using ACOLITE

Readily available surface reflectance products (Level-2) for Landsat, corrected using LaSRC, lack the quality for deriving water quality products such as Chl-*a* concentration (Tavora et al., 2023). In this chapter, Atmospheric Correction for OLI "lite" (ACOLITE, v20221114.0), developed by the Royal Belgian Institute of Natural Sciences (RBINS), has been used for converting Level-1 DNs into water-

leaving reflectance (ρ_w) values. ACOLITE provides two atmospheric correction algorithms: the default dark spectrum fitting (DSF) and exponential extrapolation (EXP) model, with the former chosen for its reliability, as EXP may overestimate errors related to adjacency effects and sunlight (Tavora et al., 2023). ACOLITE's effectiveness has been demonstrated in various Chl-*a* retrieval studies as can be seen in Table 4.1. Additionally, multiple studies have reported ACOLITE's superior performance against other processors using in-situ reflectance matchups (Pereira-Sandoval et al., 2019; Tian et al., 2022; Warren et al., 2019).

Additionally, ACOLITE was used to mask non-water pixels, including cloud, cloud shadow, land, snow, ice, or even too shallow waters where bottom reflectance might interfere. The water-leaving reflectance, ρ_w, more accurately termed water-leaving radiance reflectance (Doerffer, 2015), is defined as (Tavora et al., 2023):

$$\rho_w(\theta_v, \phi, \lambda) = \pi \frac{L_w(\theta_v, \phi, 0^+)}{E_d(\theta_s, 0^+)} = \pi R_{RS}(\theta_v, \phi, \lambda) \quad (4.1)$$

where θ_v , θ_s and ϕ are respectively viewing direction, solar zenith angle and azimuthal angle measured relative to the Sun's azimuthal direction. L_w is the water-leaving radiance just above the water surface (0^+), E_d is the downwelling irradiance, and λ is representative of each band. With the Sun at the zenith, ρ_w equals π times the nadir-viewing remote sensing reflectance R_{RS} (sr^{-1}). Both ρ_w and R_{RS} are the most used radiometric products in Chl-*a* retrieval studies.

4.2.5 Chl-*a* Retrieval Model

The Chl-*a* retrieval methods found in the literature exhibit a large diversity, reflecting varying approaches by different researchers. Mishra et al. (2017) and Olivetti et al. (2020) separated these methods into five categories: empirical, semi-empirical, analytical, semi-analytical, and quasi-analytical. In recent years, with advancements in machine learning (ML) in remote sensing, a distinct category for ML models has been recognized for describing Chl-*a* retrieval algorithms (Chegoonian et al., 2023).

In this chapter, semi-empirical approach is utilized. The semi-empirical models establish a regression-based relationship, often explicit, between in-situ Chl-*a* concentration and remote sensing reflectance of bands or band combinations that are referred to as reflectance indexes. These indexes range from simple two-band algorithms such as band ratios or normalized difference indexes (Ogashawara et al.,

2021), to three-band algorithms like KIVU, and even four-band algorithms such as Surface Algal Bloom Index (SABI) (Beck et al., 2016; Buma & Lee, 2020).

Initially, the 8 reflectance indexes described in Table 4.3 were selected as candidates for developing a semi-empirical model. A feature importance analysis, based on SHapley Additive exPlanations (SHAP) values and correlation coefficients, was conducted using the matchup concentrations and the averaged reflectance indexes within a 3 ×3 window of co-located pixels. In this context, SHAP values serve as a powerful tool to interpret the predictive model by assigning an importance score to each input feature, revealing how significantly each reflectance index influences the estimation of Chl-*a* concentration (Appling et al., 2022). The feature importance analysis identified the blue-to-red ratio index (I_3) as particularly significant (see Figure A2.4).

Table 4.3. Selected Chl-*a* retrieval indexes as input features of predictive models

Index Group	**Index Annotation**	**Band Arithmetic**	
Band Ratios	I_1	$\frac{UB}{B}$	$\frac{\rho(443)}{\rho(482)}$
	I_2	$\frac{B}{G}$	$\frac{\rho(482)}{\rho(562)}$
	$\boldsymbol{I_3}$	$\boldsymbol{\frac{B}{R}}$	$\boldsymbol{\frac{\rho(482)}{\rho(655)}}$
	I_4	$\frac{G}{R}$	$\frac{\rho(562)}{\rho(655)}$
	I_5	$\frac{NIR}{R}$	$\frac{\rho(865)}{\rho(655)}$
Normalized Difference Vegetation Index	I_6	$\frac{NIR-R}{NIR+R}$	$\frac{\rho(865)-\rho(655)}{\rho(865)+\rho(655)}$
FLH-based Indexes	I_7	$G-UB-(\frac{\lambda_G-\lambda_{UB}}{\lambda_R-\lambda_{UB}})(R-UB)$	$\rho(562)-\rho(443)-\left(\frac{562-443}{655-443}\right)[\rho(655)-\rho(443)]$
	I_8	$G-R-(\frac{\lambda_G-\lambda_R}{\lambda_B-\lambda_R})(B-R)$	$\rho(562)-\rho(655)-\left(\frac{562-655}{482-655}\right)[\rho(482)-\rho(655)]$

The selection of the blue-to-red ratio index for estimating Chl-*a* concentration is supported for several reasons. Firstly, it is particularly effective for estimating Ch-*a* in turbid waters (Hafeez et al., 2019; Paltsev & Creed, 2022; Sherman et al., 2023; Stumpf et al., 2016) where the optical properties of non-algal particles significantly affect reflectance, which is characteristic of case two waters like the hotspots at river outflows of Western Lake Ontario (Salem et al., 2017). Secondly, the index has been recommended for its efficacy in detecting cyanobacteria (Shi, Zhang, Qin, et al., 2019; Sun et al., 2015), prevalent in environments such as Hamilton Harbour. Lastly, numerous studies have validated the use of this ratio for Chl-*a* retrieval from Landsat imagery (Han & Jordan, 2005; Matus-Hernández et al., 2018; Sass et al., 2007; X. Yang et al., 2020), underscoring its applicability and reliability across

different water bodies and conditions. Several regression types, including linear, quadratic, cubic, and exponential, were tested with the blue-to-red ratio as input. The exponential regression appeared as the most promising fit (see Figure A2.5).

4.2.6 Performance Metrics

In this chapter, based on the work of Seegers et al. (2018) and more recent recommendations by Pahlevan et al. (2020b), the coefficient of determination (R^2), was replaced by more informative metrics such as RMSE, MAE and MAPE. The assumption of a Gaussian distribution without outliers underlying the use of R^2 is unlikely to be met in water colour remote sensing with limited matchup counts (Seegers et al., 2021). RMSE measures the standard deviation of prediction errors, MAE averages the absolute errors, and MAPE expresses the error as a percentage of actual values. Although RMSE is still sensitive to the distribution of errors and assumes a Gaussian distribution, it is a common statistical metric and widely used in the literature. MAE and MAPE on the other hand provide a more straightforward error quantification. MAE is particularly effective as it does not exaggerate outliers and accurately represents error magnitude. The equations for these metrics are:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - E_i)^2} \quad (4.2)$$

$$MAE = \left(\frac{1}{n}\sum_{i=1}^{n}|M_i - E_i|\right) \quad (4.3)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left(\frac{M_i - E_i}{M_i}\right) \quad (4.4)$$

where n, M and E represents the number of samples, the measured (in-situ) and estimated (modeled) Chl-*a* concentration, respectively.

4.2.7 Algal Bloom Indicators

In the literature, different approaches are used to identify bloom pixels based on satellite-derived Chl-*a* concentration products. Studies like those by Cao et al. (2020), and Ma et al. (2022), used a threshold on the Floating Algae Index (FAI) to flag algal blooms. Others, such as Tan et al. (2017), applied a Chl-*a* concentration threshold of ≥ 20 µg/L, whereas Binding et al. (2021, 2023), and Zeng & Binding (2021) use a more conservative threshold of ≥ 10 µg/L. This lower threshold aligns with the World Health Organization's guideline for bloom conditions in recreational waters, indicating relatively mild

or low risk of adverse health effects in cyanobacteria-dominated blooms (Binding et al., 2023; Soontiens et al., 2019). In this chapter, a Chl-*a* concentration threshold of ≥ 10 µg/L is selected to flag bloom pixels.

Next, various indicators for analyzing algal bloom conditions from satellite-derived Chl-*a* estimates were evaluated. Commonly used algal bloom indicators include bloom intensity, bloom extent (or bloom area), bloom severity, and bloom duration, with the latter often used in time-series studies that use sensors with shorter revisit times (Feng et al., 2021; Ho et al., 2019; Ma et al., 2022; Soontiens et al., 2019; Yin et al., 2023; Zeng & Binding, 2021). Following the work of Binding et al. (2018, 2021, 2023), the selected algal bloom indicators for this study are: (i) bloom intensity (µg/L) as the average Chl-*a* concentration of flagged pixels, (ii) bloom extent (km²) as the total area covered by flagged pixels, and (iii) bloom severity (µg km²/L) as the product of the first two indicators. The indicators produce three quantitative metrics for each Chl-*a* concentration map (that is, each scene) that, when combined with other outputs such as annually and monthly averaged estimated Chl-*a* concentration maps plus trend analyses, for example based on the Mann-Kendall test, provide a general view of bloom conditions across the study area and their intra- and inter-annual variability and multi-year trends, as well as the detection of anomalies and bloom hotspots.

4.3 Results

4.3.1 Time-Series Plots of Algal Bloom Indicators

After calibrating and validating the retrieval model with 5-fold cross-validation, using an 80/20 % split for training and testing respectively, the model was applied to the entire preprocessed Landsat 8 and 9 collections, to generate estimated Chl-*a* concentration maps. Subsequently, three algal bloom indicators, namely, bloom intensity (Figure 4.4), bloom extent (Figure 4.5), and bloom severity (Figure 4.6), were calculated for each scene. These indicators were plotted on time series figures for Hamilton Harbour, as well as the nearshore and offshore waters of Western Lake Ontario. A perfect (lossless) image would encompass approximately 23000, 668000, and 6600000 pixels for Hamilton Harbour, nearshore, and offshore areas, respectively, at a 30×30 m^2 spatial resolution. However, factors such as partial swaths, cloud cover, lake ice, and various quality masks reduce the number of usable pixels per image. To address data reliability issues, a fourth indicator representing the percentage of data availability for each scene was introduced, visualized as vertical lines behind each scatter point, with

availability ranging from 0 to 100 percent. This metric helps in assessing coverage extent for each area of interest per scene, where a darker shade signifying 100% availability, representing a perfect overview of the area with no pixel loss. Annual and monthly averages for each time series are displayed on the right side of each plot, highlighting the annual trends, anomalies, and seasonality of algal blooms. Notably, calculations for each indicator are derived solely from flagged pixels. The grey and black bars on the plots represent the average and maximum values of each indicator, respectively. It is worth noting that the maximum bars depict the peak average of each pixels with ≥ 10 µg/L estimated Chl-*a* concentration for a given year or month, rather than the maximum pixel value (estimated Chl-*a* concentration) within each scene.

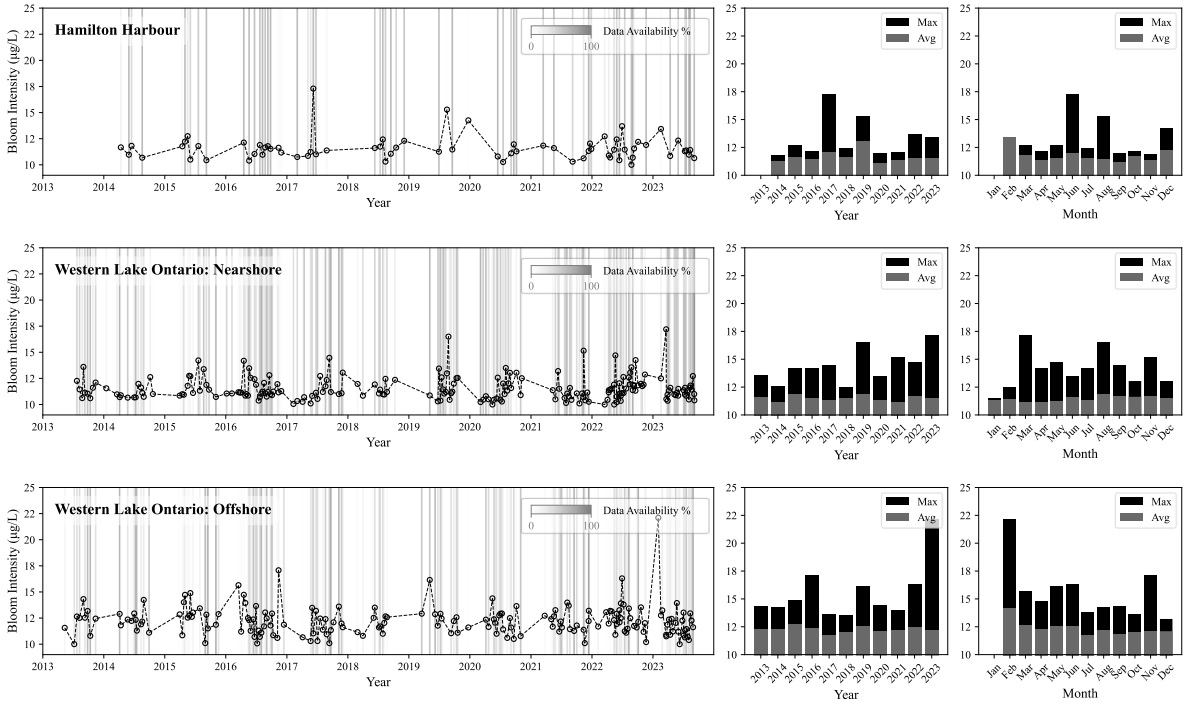

Figure 4.4. Time-series plots of bloom intensity for Hamilton Harbour, nearshore and offshore waters of Western Lake Ontario. Scatter plots on the left represent the temporal dynamics of bloom intensity from 2013 to 2023 where the background bars depict data availability for each scene. The bar plots on the right represent annual and monthly bloom intensity averages and maximums.

In Hamilton Harbour, bloom intensity ranges often between 10.0 to 12.0 µg/L with spikes up to 18.0 µg/L in June 2017. Nearshore regions exhibit slightly lower intensities, with averages around 11.0 µg/L and the peak of 17.5 µg/L in March 2023. Offshore intensities vary between 10.5 to 13.5 µg/L, with a significant spike of 21.5 µg/L in February 2023. The annual averages of bloom intensity in Hamilton

Harbour ranges from 11.0 µg/L in 2020 to 12.5 µg/L in 2019, with monthly averages range between 11.0 µg/L (September) to 13.0 µg/L (February), excluding 2013 and January due to lack of data. Offshore regions' overall bloom intensity averages are marginally higher by 0.5 µg/L compared to nearshore, without notable annual trends, though a decline after the peaks in 2019 and an uptrend in maximums thereafter are observed. Nearshore regions show an increasing bloom intensity monthly trend from January to August, followed by a downtrend back to January. Hamilton Harbour has seen diminishing peaks since 2017, whereas Lake Ontario reached its highest eleven-year records in 2023.

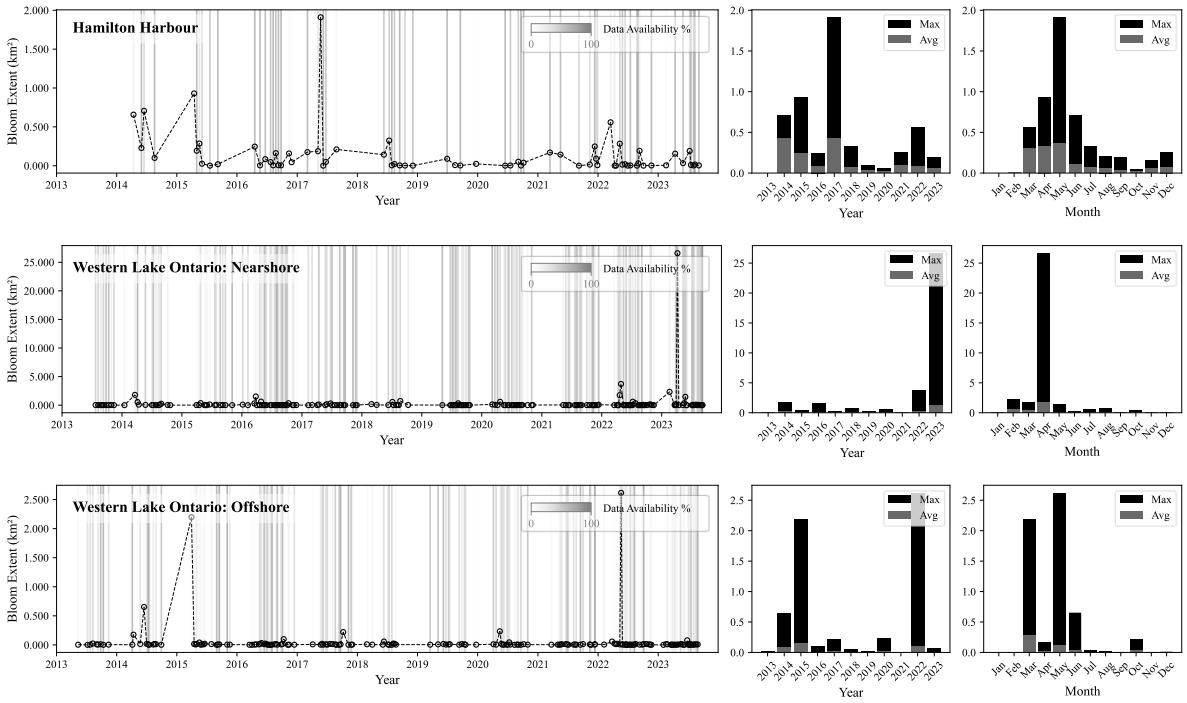

Figure 4.5. Time-series plots of bloom extent for Hamilton Harbour, nearshore and offshore waters of Western Lake Ontario. Scatter plots on the left represent the temporal dynamics of bloom extent from 2013 to 2023 where the background bars depict data availability for each scene. The bar plots on the right represent annual and monthly bloom extent averages and maximums.

Figure 4.5 and Figure 4.6 depict bloom extent and severity respectively. Despite its size, Hamilton Harbour often records higher bloom extents due to its eutrophic to hypereutrophic conditions, contrasting with the oligo-mesotrophic nearshore and oligotrophic offshore waters of Western Lake Ontario, with extents up to 1.8 km², about 10% of its area. Notably, anomalies include a 25.5 km² plume in April 2023 from southern hotspots and significant events in offshore waters in May 2022 (2.5 km²) and March 2015 (2.2 km²), indicating substantial bloom events. However, the majority of measured

extents in Western Lake Ontario are below 1.0 km², suggesting localized blooms. The subsequent maps will provide further hotspot insights. Trend analysis for Hamilton Harbour indicates a decreasing bloom extend trend, except for, a peak in May 2017. Its monthly distribution shows an uptrend from October to May, peaking from March to September, and then declining. Nearshore and offshore areas show no consistent trends, with only extreme events notably affecting bloom extent measurements. However, offshore maximums show growth similar to their intensity over the past eleven years.

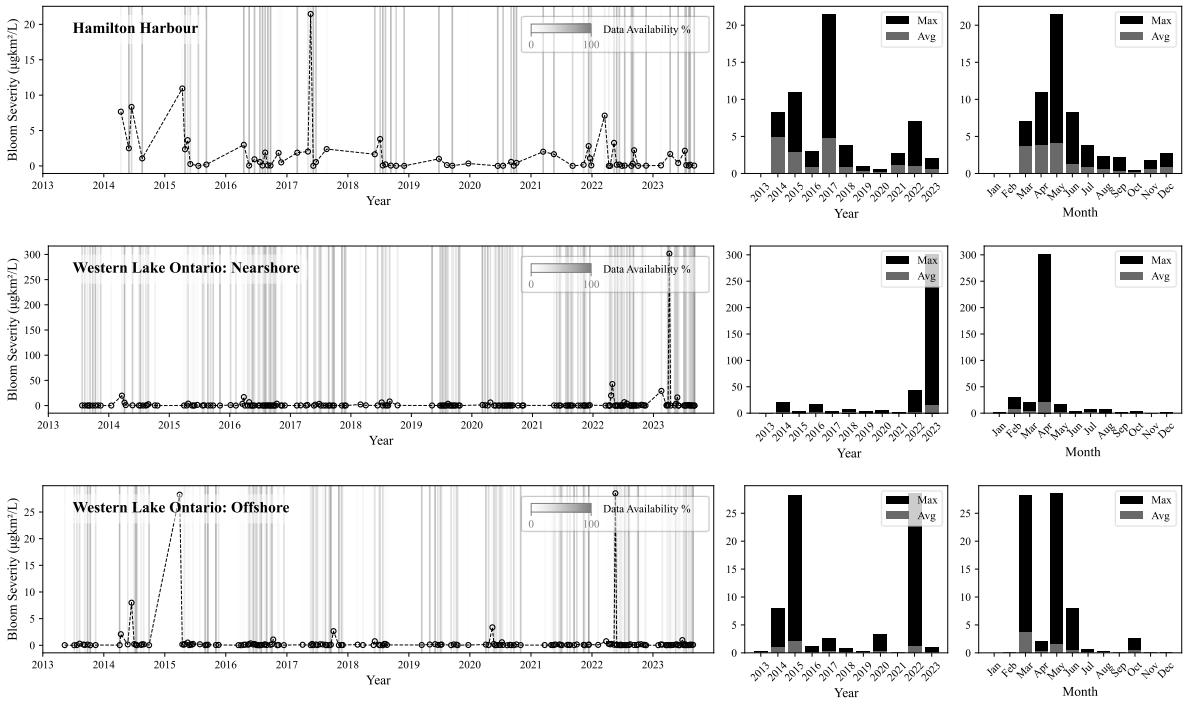

Figure 4.6. Time-series plots of bloom severity for Hamilton Harbour, nearshore and offshore waters of Western Lake Ontario. Scatter plots on the left represent the temporal dynamics of bloom severity from 2013 to 2023 where the background bars depict data availability for each scene. The bar plots on the right represent annual and monthly bloom severity averages and maximums.

Bloom severity, as illustrated in Figure 4.6, is defined as the product of intensity and extent, usually remains low in nearshore and offshore areas, except during significant events. In Hamilton Harbour, severity ranges from 0 to 4, peaking at 22 µg. km²/L in May 2017. Yearly and monthly severity trends closely follow those of bloom extent.

4.3.2 Averaged Annual and Monthly Maps of Estimated Chl-*a* Concentration

In addition to calculating bloom indicators, averaged yearly and monthly maps were generated to analyze trends, seasonality, and hotspots. Unlike the three indicators based on flagged pixels with ≥ 10 µg/L Chl-*a* concentration, these maps average across all available pixels grouped by year or month. Therefore, while related, they serve distinct purposes: the indicators quantify intensity and extent, and the maps visually represent bloom condition over time.

Figure 4.7 displays the estimated Chl-*a* concentration, averaged across all images within specific years. It includes a 10 µg/L algal bloom threshold and Carlson's trophic state index (TSI) thresholds (Carlson, 1977). Non-water areas are marked white, indicating land or clouds. Notably, in 2013 and 2018, image availability was lowest, leading to significant pixel loss in the Hamilton Harbour and central lake areas due to cloud cover. The launch of Landsat 9 in 2021 significantly increased image availability, offering a clearer view of the area in recent years.

The averaged annual maps reveal few eutrophic areas except for specific nearshore (like St. Catharine to Niagara-on-the-Lake waterfront, notable in 2023) and northern hotspots, like Toronto Harbour in 2022 or Frenchman's Bay in 2014. Hamilton Harbour consistently shows more pronounced eutrophic conditions, particularly in its eastern parts near Royal Botanical Gardens (RBG), illustrated as yellow (eutrophic) in most of the years. These maps offer a broad perspective for assessing yearly changes and support the analysis from the indicators. For example, Hamilton Harbour's lowest bloom intensity and extent in 2020 correlate with bluer hue in that year's map, indicating reduced algal activity compared to greener hues in other years. Figure 4.8 presents monthly averages for estimated Chl-*a* concentration maps, better suited for identifying seasonal patterns and pinpointing season-constrained hotspots.

Regarding the seasonal variations, for Hamilton Harbour, the lowest bloom occurs from October to February, consistent with indicator trends. The monthly maps are valuable for highlighting April to May as more eutrophic, supplementing indicator data by showing nuanced seasonal shifts. Overall, for Hamilton Harbour, as observed in the averaged monthly maps, we can identify March to September as critical periods for algal bloom monitoring. A similar period, but with milder concentrations, also applies to the algal bloom of the northern shore hotspots, such as Frenchman's Bay. For the southern shoreline, the period from November to April is marked by lighter pixel colours. However, since this period falls during the winter months, it's important to take into account the high input of sediments from the Niagara River, creating a plume which flows easterly towards US shores. This increased turbidity interferes with the optical properties of water, thereby making Chl-*a* estimates for this area during this period, less reliable. Furthermore, similar to the previous figure, cloud cover significantly

affected both image and pixel availability in December and January leaving some missing pixels in the maps. During 2013-2023, only 17 images acquired in December were available for processing, compared to 47 in August, indicating clearer skies in the latter month. Further in-depth interpretation of these results will be discussed in the following section.

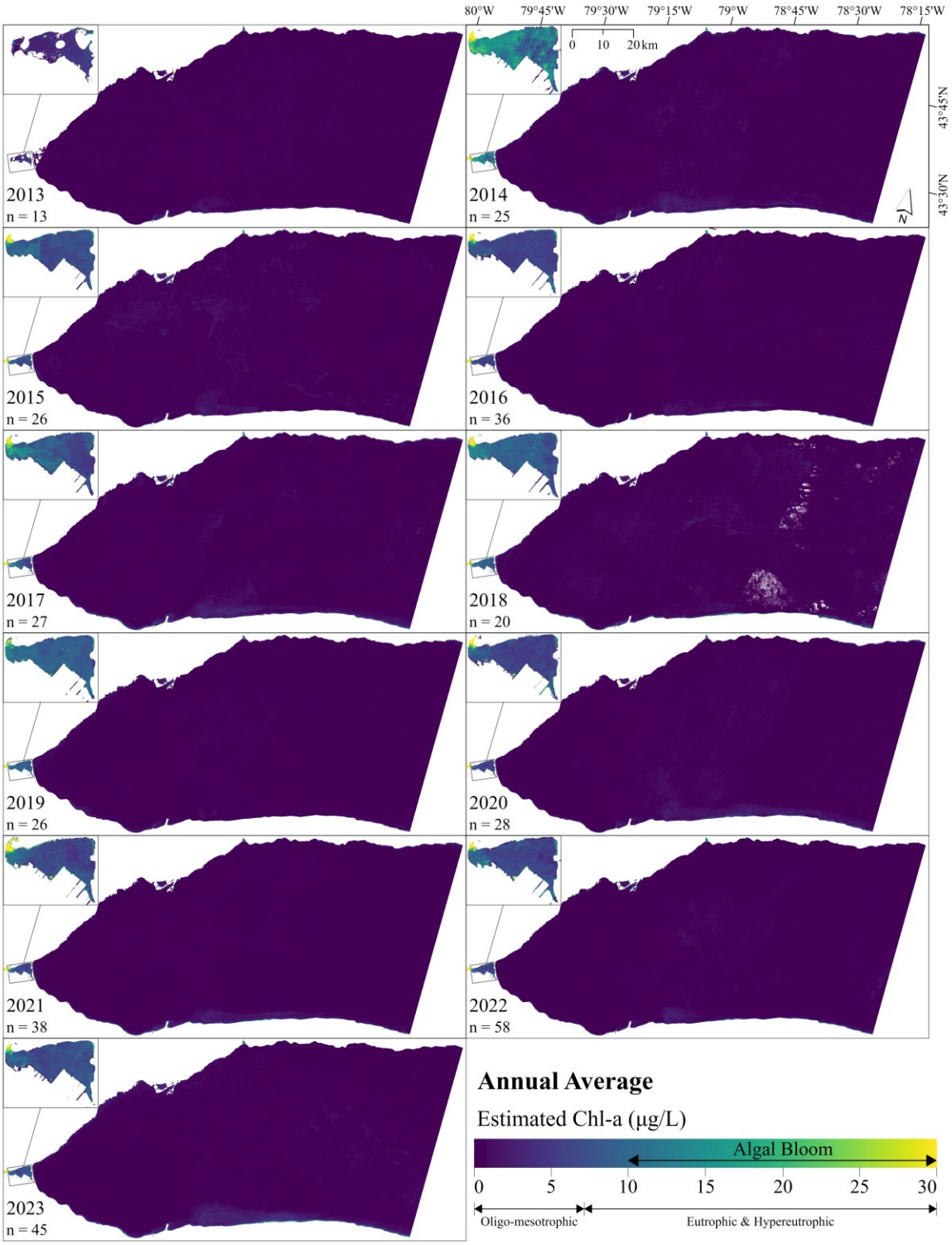

Figure 4.7. Annually averaged maps of estimated Chl-*a* concentration for Hamilton Harbour and Western Lake Ontario, derived from Landsat 8 and 9 imagery acquired between 2013 and 2023

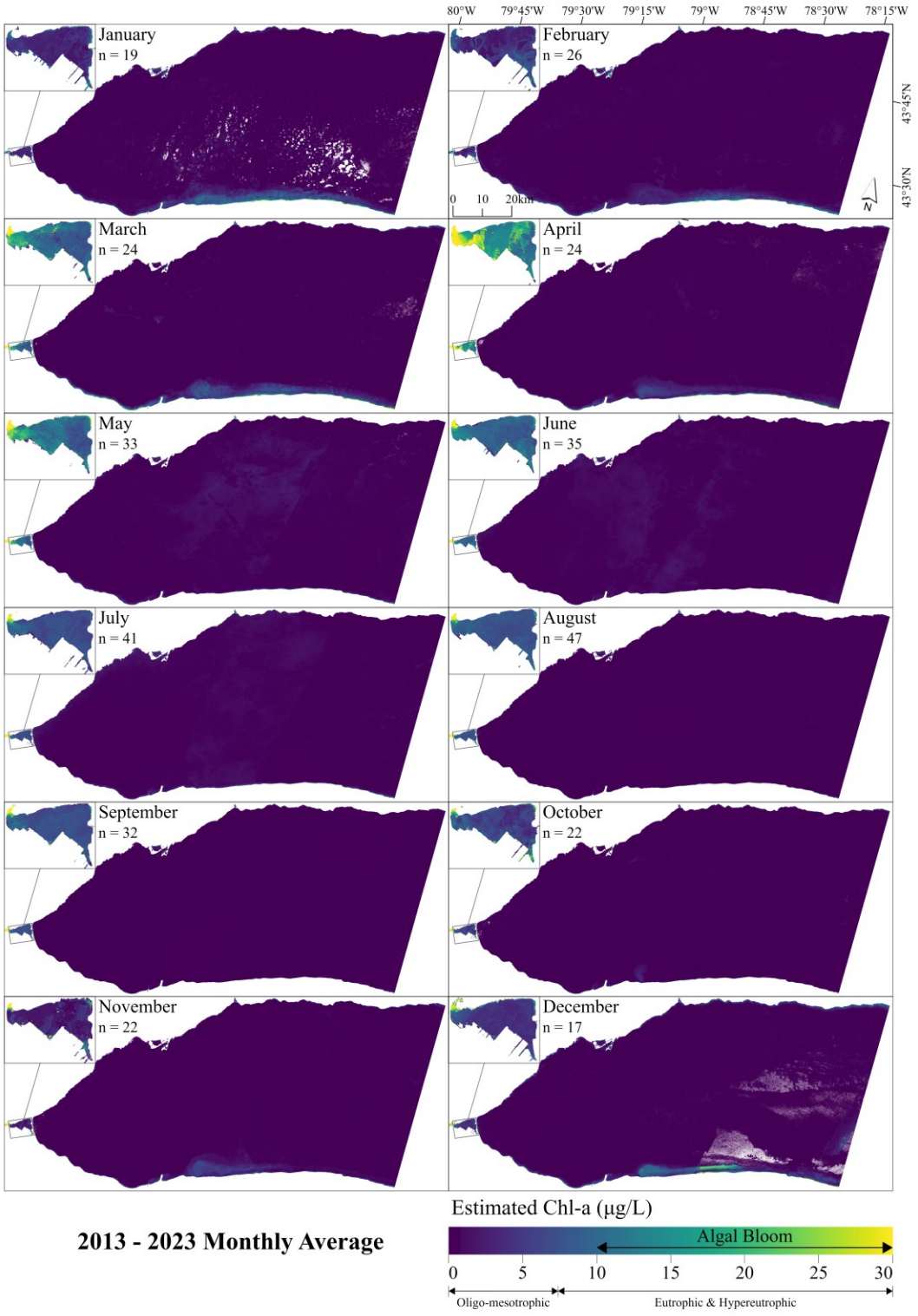

Figure 4.8. Eleven-year monthly averaged maps of estimated Chl-*a* concentration for Hamilton Harbour and Western Lake Ontario, derived from Landsat 8 and 9 imagery acquired between 2013 and 2023

4.4 Discussion

While single bloom indicators offer valuable insights, they might not fully capture the complexity of bloom conditions. A bloom could be extensive yet have low biomass or small and short-lived but have high biomass (Binding et al., 2021). Moreover, elevated Chl-*a* concentration suggest potential scum formation without necessarily indicating toxin presence. This section uses previously defined indicators and yearly/monthly maps to analyze algal blooms' patterns and variations in Hamilton Harbour and Western Lake Ontario from 2013 to 2023, as well as addressing limitations and uncertainties.

4.4.1 Spatial-Temporal Variability and Trends

Comparing our bloom intensity results with in-situ records of Hamilton Harbour reveals general underestimation by our metrics (Figure A2.6), especially for years with extreme outliers like years 2014, 2017, and 2021. However, the alignment within a ±1 µg/L margin of error for 2013 to 2016 indicates reasonable accuracy. Monthly trend analysis shows an underestimate for late summer blooms, despite earlier onset of blooms observed, starting as early as March and peaking in April, differing from past decades' data indicating a summer peak as shown by Hiriart-Baer et al. (2009) that indicated summer (June-August with 14.1±8.6 µg/L) Chl-*a* averages nearly twice those of spring (March-May with 7.6±6.6 µg/L). Our model successfully captured the 2017 peak but failed to predict the 2014's extreme outliers, with records as high as 185 µg/L, indicating limitations in handling such anomalies. Limited in-situ data in 2019 and 2021 due to COVID-19 restrictions, and the absent of publicly released data for 2022 and 2023 highlight the advantages of remote sensing for continuity in environmental monitoring when field activities are restricted.

Observations reveal no significant trend in bloom intensity and severity for Lake Ontario's nearshore and offshore waters. A comparison of our annual average estimated Chl-*a* concentration with MODIS-derived lake-wide annual average values reported on NOAA's National Marine Ecosystem Status website (NOAA, 2024) indicates high correlation but a steady trend, likely suggesting stable external nutrient loading over the past decade. The NOAA's satellite-derived estimates are based on a band-ratio retrieval algorithm developed from Great Lakes National Program Office (GLNPO) monitoring data (Lesht et al., 2013, 2016), applied to MODIS images from 2013 to 2020. The results of this comparison, presented in Figure A2.7, with further discussion in the supplementary materials, show that both estimates are highly correlated, covering the range of 1.5 to 2.0, consistent with in-situ measurements (Estepp & Reavie, 2015). A slight decrease in bloom extent and severity for Hamilton Harbour may

reflect localized, concentrated algal blooms, potentially influenced by circulation patterns within the harbour. Annual trend analysis across averaged maps shows an increasing bloom trend in Hamilton Harbour from 2014, peaking in 2018, decreasing until 2020, and then rising again up to 2023. Monthly trends suggest higher bloom intensity from March to September, consistent with in-situ data, highlighting the importance of focused monitoring during these months.

Nearshore areas demonstrate a bell-shaped bloom intensity trend, with a peak in August, aligning with field-based observations and indicating nutrient-rich conditions favourable for algal proliferation at river mouths. Notably, the Niagara plume poses a significant algal bloom risk due to its turbidity and phosphorus levels. Frenchman's Bay shows a slight decrease in estimated Chl-*a* concentration, aligning with other studies indicating offshore oligotrophic conditions and a general decrease in water quality parameters from the shoreline to offshore.

4.4.2 Uncertainties and Limitations

Time-series analysis of satellite-derived Chl-*a* concentration faces several challenges and uncertainties. Despite remote sensing ability to provide broad views of HABs in Western Lake Ontario, understanding the technology's limitations and the uncertainties in derived products is essential. The key challenge is the lack of frequent, long-term in-situ and remote sensing matchups during the model calibration/validation. Finer temporal resolution satellite data often compromise spatial resolution for shorter revisit times (C. Hu et al., 2019; Ma et al., 2022), affecting the use of Chl-*a* concentration maps for small areas like Hamilton Harbour. Spectral resolution limitations, especially OLI's suboptimal settings for turbid waters and the absent of red-edge (690-790 nm) band, complicate analysis (Cao, Ma, Melack, et al., 2022). Additional challenges include adjacency effects, affecting pixels up to 40 km away from the shoreline, the stripping effects, sunglint (Tavora et al., 2023), and the interference of atmospheric columns as well as lakebed (bottom reflectance), necessitating complex correction methods.

In-situ data uncertainties arise from measurement depth variability, and intrinsic error in laboratory extraction techniques, such as High Pressure Liquid Chromatography (HPLC), fluorometry, and spectrophotometry, with the latter having error margins up to 70% (Chegoonian et al., 2023; Seegers et al., 2021; Smith et al., 2018). The accuracy of Chl-*a* retrieval is also affected by temporal difference between in-situ measurements and satellite overpasses (in this case ±2 days) and the representation of in-situ samples within larger pixel windows (in this case 3×3 equal to 90×90 m^2), and spatial patchiness

of the bloom within them (Seegers et al., 2021; Soontiens et al., 2019), complicating satellite-derived bloom analysis.

The selection threshold for algal bloom flagged pixels and bloom indicators further complicates results interpretation and inter-study comparisons. The regression relationship developed for the Chl-*a* model tends to saturate at around 30 μg/L underestimating values at extremely high concentrations. Furthermore, the impact of other optically active water constituents (OACs) significantly alters the water-leaving reflectance, challenging semi-empirical band-ratio algorithms in optically complex waters (Warren et al., 2021). These algorithms are often confounded by the presence of suspended sediments and coloured dissolved organic matter (CDOM) (Vélez-Nicolás et al., 2021), which dominate the optical properties and satellite-derived reflectance in blue-green wavelengths (Le et al., 2013). This introduces uncertainties, particularly in low Chl-*a* concentration scenarios. Additionally, the variability of local inherent optical properties (IOPs) due to the diversity in phytoplankton community compositions, introduces additional bias (Binding et al., 2021).

Moreover, bloom severity indicators do not guarantee corresponding risks of bloom toxicity, due to variations in phytoplankton community composition and toxigenicity. The absence of a unique optical signature for cyanobacterial toxins necessitates indirect, proxy-based methods for assessing bloom toxicity from satellite observations. Despite these challenges, advancements in retrieval algorithms, including optical water type classifications, ML techniques, and hyperspectral imaging, show promise in reducing uncertainties and improving accuracy of satellite-derived Chl-*a* concentration products.

4.5 Conclusions

The analysis of algal blooms in Lake Ontario, particularly through satellite remote sensing and field surveys, highlights the complexities and challenges in monitoring, understanding, and managing these ecological phenomena. The application of remote sensing, and particularly the development and use of a semi-empirical Chl-*a* retrieval model in this chapter, has provided unparalleled insights into the dynamics of algal blooms within Hamilton Harbour and Western Lake Ontario from 2013 to 2023. By utilizing Landsat 8 and 9 imagery to calculate algal bloom intensity, extent, and severity, alongside satellite-derived spatial-temporal maps, this chapter enhances our understanding of bloom conditions and the underlying drivers in the region.

In the past eleven years, Hamilton Harbour displayed a declining trend in bloom peaks, contrasting with Lake Ontario's nearshore regions, which exhibited an increasing monthly bloom intensity from January to August, before declining back to January levels. Remarkably, 2023 marked the highest bloom intensity records across Lake Ontario, emphasizing the variability in bloom dynamics. Hamilton Harbour often reported higher bloom extents, attributed to its eutrophic to hypereutrophic conditions, in stark contrast to the predominantly oligotrophic conditions of the lake's offshore waters. This resulted in most bloom extents in Western Lake Ontario being localized and below 1.0 km^2. Notably, the spatial analysis underscored few eutrophic areas aside from specific nearshore and northern hotspots, with Hamilton Harbour consistently showcasing pronounced eutrophic conditions near the Royal Botanical Gardens. Seasonal variations revealed the lowest blooms in Hamilton Harbour occurring from October to February, with an earlier onset observed in March, peaking in April. Despite these localized and temporal variations, Lake Ontario's nearshore and offshore areas showed no significant trend in bloom intensity and severity, suggesting a decade of stable external nutrient inputs. The monthly trends supported the necessity for heightened monitoring from March to September, particularly in nearshore areas where a bell-shaped bloom intensity trend peaks in August, aligning with field observations and emphasizing the nuanced interplay of ecological dynamics across these regions.

These collective findings enhance our comprehension of HABs in Lake Ontario, highlighting the significant roles of ecological, phenological, and anthropogenic factors in bloom dynamics. By providing detailed spatial-temporal maps and indicators, remote sensing not only advances our understanding of algal bloom dynamics but also supports the development of more informed and dynamic management strategies. This holistic approach, which merges traditional field surveys with cutting-edge remote sensing technologies and predictive modeling, makes a significant advancement in our efforts to mitigate impacts of HABs and ensure the sustainability of aquatic ecosystems in the Great Lakes region. Such integration is crucial for developing comprehensive strategies that address the multifaceted nature of HABs, thereby contributing to the protection and preservation of these vital water bodies in the face of evolving environmental challenges.

Chapter 5
Conclusions and Perspectives

5.1 Summary of Key Findings

This thesis presents a comprehensive exploration of Chl-*a* concentration retrieval and the monitoring of algal abundance in Western Lake Ontario and Hamilton Harbour using space-borne remote sensing spanning multiple satellites. The key findings from the study are summarized as follows.

The performance of 27 commonly used Chl-*a* retrieval indexes across eleven types of Level-2 products derived from four distinct satellite sensors was analyzed based on a matchup analysis comprising 600 in-situ surface water Chl-*a* data points collected from 2000 to 2022. The study identified that the effectiveness of retrieval is significantly influenced by several factors, including the choice of satellite platform, in-situ Chl-*a* data categorization (especially according to the range of concentrations), atmospheric correction processes, and the specific Chl-*a* concentration retrieval indexes employed. ACOLITE-corrected Landsat 8 and Sentinel-2 products coupled with simple two band ratios of, for instance, blue-to-green or blue-to-red, were highlighted as one of the superior schemes for Chl-*a* concentration retrieval, especially for eutrophic and hypereutrophic waters. These findings laid the groundwork for the methodology development in the subsequent time-series analysis and, with caution, could be generalized for other large water bodies with similar characteristics.

Through the development and application of a semi-empirical Chl-*a* retrieval model, Chapter 4 yielded Chl-*a* concentration maps and algal bloom indicators using Landsat 8 and 9 imagery acquired between 2013 and 2023. For Western Lake Ontario, the results clearly identified nearshore areas receiving stream or channel discharge as hotspots for algal proliferation. Specifically, Chapter 4 examined the plume formed near the mouths of the Niagara River and Welland Canal that supply relatively phosphorus-rich waters from Lake Erie to Lake Ontario. This enabled us to delineate the plume's extension along the eastern US shoreline of Lake Ontario. Additionally, the trends and seasonality of several northern hotspots along the Mississauga-Pickering waterfront, including Toronto Harbour and coastal wetlands like Frenchman's Bay, were mapped. Overall, no significant trends in algal growth intensity or severity were observed for Lake Ontario's nearshore and offshore waters in the last eleven years, likely implying stable nutrient loading over the past decade. However, a slight decreasing trend in bloom extent and severity was observed for Hamilton Harbour with localized bloom occurrences in specific areas like Bayfront Park Beach due to circulation patterns. The study also highlighted the potential of remote sensing for near-real-time Chl-*a* monitoring across large areas but also for filling

gaps in historical data, such as during the 2020 COVID shutdowns. The results of the data series analyses further suggest a trend of algal growth appearing earlier than the historical average conditions, hence, stressing the importance of extending monitoring programs beyond the traditional time window in order to detect early blooms starting in March.

Another finding from this study is that increased model complexity does not necessarily lead to better Chl-*a* retrieval accuracy, advocating for balancing model simplicity and effectiveness in remote sensing applications for water quality monitoring. (Note: in a separate study not reported in this thesis, a similar conclusion for high-spatial-resolution Chl-*a* mapping using drones was made.) Overall, the message of this thesis is that remote sensing (both satellite and drone based) can complement and augment traditional water quality monitoring of lakes. Such integration would benefit the design and assessment of environmental management and mitigation strategies that address eutrophication issues. Taken together, the thesis contributed to the fields of aquatic remote sensing by demonstrating the value of using satellite data for Chl-*a* concentration retrieval and algal bloom monitoring.

5.2 Research Challenges

This section acknowledges several challenges inherent to the methodologies and technologies of remote sensing of water quality monitoring and algal bloom detection, employed across the various chapters.

A key challenge highlighted in Chapter 3 is the highly variable performance of different satellite platforms, atmospheric correction processors and retrieval indexes, which is significantly influenced by the quality – extraction accuracy and range – and, quantity – spatial and temporal distribution and count – of matchup data used in training and testing the retrieval models. This emphasizes the critical need for sufficient in-situ data to develop reliable models and to provide improved overlap in temporal windows between in-situ data and with satellite image acquisition where the latter can be severely reduced in regions with extensive cloud cover.

Chapter 4 discusses the inevitable trade-off between temporal and spatial resolution in the selection of satellite data for time-series environmental studies. High-spatial-resolution sensors are preferred for local case studies, especially for turbid waters or shoreline waters where land and water pixels may overlap, and lakebed reflectance can interfere with the sensed signals. However, the drawback of using high-spatial-resolution sensors, such as Landsat 8 and Sentinel-2, is their longer revisit times, making time-series analysis challenging, especially for analyzing bloom durations. Additionally, the high cloud

coverage in our case study significantly reduces the number of available images, particularly during wet months. Despite these limitations, high-spatial-resolution satellites are still preferred over medium-spatial-resolution (< 1000 m) sensors with near-daily coverage when considering nearshore regions. The latter would be inadequately covered by, for instance, the 300-m resolution Sentinel-3, even though it might provide a more comprehensive set of imagery and potentially more reliable models due to its optimal spectral resolutions and additional bands dedicated to Chl-*a* retrieval.

In conclusion, while these challenges present significant limitations, potential solutions exist, such as developing hybrid models that combine data from multiple satellite platforms to leverage the strengths of each, as well as ramping up in-situ data collection efforts to support the remote sensing model calibration and validation.

5.3 Research Applications

This section discusses several potential real-world applications, elaborating on how the research findings can be utilized in practical scenarios or benefit related stakeholders.

The insights from Chapter 3 could pave the way for future research in developing robust methodologies by selecting the appropriate retrieval scheme for a given specific case study. Future machine learning approaches could build on the feature importance analysis and the identification of effective spectral indexes for Chl-*a* retrieval, as well as atmospheric correction methods across different scenarios. Such approaches could create more accurate and efficient models that focus on the important features identified in this research.

The results of Chapter 4, especially the satellite-derived indicators for bloom intensity and extent, could prove valuable for local conservation authorities and environmental agencies. These indicators enable the identification of algal bloom hotspots and periods of heightened risk, facilitating efficient resource allocation, targeted monitoring, and timely mitigation measures to minimize the socio-economic and ecological impacts of harmful algal blooms in the region. Additionally, the findings from this chapter fill in spatial and temporal data gaps, providing a clearer picture of periods (such as winter months or during the 2020 COVID-19 shutdown) when no field sampling is conducted. Expanding on this, additional usages of satellite-derived Chl-*a* products based on near-real-time acquired images – assuming the models are trained and tested in advance with sufficient matchup data – can be considered. For instance, web applications, say, within Esri's ArcGIS StoryMaps, Google Earth Engine (GEE)

environment, or standalone web applications such as EOLakeWatch by ECCC or CyAN by US EPA, can be leveraged to incorporate near-real-time water quality products for any region of interest and accessible to the public, to check the status of beaches, for example, while decision-makers could use them as early-warning tools for planning in-situ inspections.

Overall, the applications of this research extend beyond the academic realm, offering possibilities for practical tools and methodologies to enhance the monitoring of algal dynamics. By integrating space-borne remote sensing technologies with conventional environmental monitoring approaches, stakeholders across various sectors can achieve more informed, effective, and sustainable outcomes in water quality management.

5.4 Recommendations for Future Research

In this section, we provide recommendations for future research, some of which complement the findings of each chapter to improve results, while others could constitute the basis for separate studies.

- Building on the work in Chapter 3 using semi-empirical models, we can envision conducting a comprehensive comparison study on the same 600 matchup dataset with various machine learning models, including support vector regression (SVR), random forest regression (RFR), extreme gradient boosting (XGBoost), and more innovative neural network models like mixture density network (MDN).
- Currently, only a very limited number of atmospheric correction processors, such as SIAC, are available within the GEE environment for converting Level-1 images to Level-2 products on the backend, and their validity has not been tested as extensively as other commonly used processors. Assuming that more processors become available for use in GEE in the future, one could develop a web application for converting Level-1 products to the desired Level-2 and then Level-2 to Chl-*a* concentration maps, with no processing load on the client's hardware. A similar analysis to that in Chapter 4 could also be programmed, and algal bloom indicators could be calculated and visualized with just a few clicks from the user. Of course, given GEE's current processing limitations, there would be a trade-off between the temporal coverage and its spatial extent of such time-series analysis. However, for an area as large as Hamilton Harbour, this proposal is feasible.

- For more reliable results on the seasonality of algal blooms in the region, it is recommended to add resampled Sentinel-2 images through Harmonized Landsat and Sentinel-2 (HLS) for a period after the launch of Landsat 9 in September 2021 (for example, September 2021 to September 2023). This approach would reduce the revisit times to 2 days and significantly increase the number of available matchups through using four satellites instead of just the two constellations of Landsat 8 and 9 or Senintel-2 A/B separately. Such a collection, specifically for studying the seasonality of blooms and not for time-series analysis, is recommended for regions where sufficient in-situ sample collection is available.

All these recommendations are contingent upon the availability of sufficient budget and logistical support, which are critical for the successful execution and completion of these projects. Each of the recommendations aims to push the boundaries of current methodologies and technologies applied in this research, fostering innovation, and improving the understanding and management of water quality and ecosystem health by means of remote sensing.

References

Abdelal, Q., Assaf, M. N., Al-Rawabdeh, A., Arabasi, S., & Rawashdeh, N. A. (2022). Assessment of Sentinel-2 and Landsat-8 OLI for Small-Scale Inland Water Quality Modeling and Monitoring Based on Handheld Hyperspectral Ground Truthing. *Journal of Sensors*, *2022*, 1–19. https://doi.org/10.1155/2022/4643924

Adams, H., Ye, J., Persaud, B. D., Slowinski, S., Kheyrollah Pour, H., & Van Cappellen, P. (2022). Rates and timing of chlorophyll- a increases and related environmental variables in global temperate and cold-temperate lakes. *Earth System Science Data*, *14*(11), 5139–5156. https://doi.org/10.5194/essd-14-5139-2022

Ahamed, T., Tian, L., Zhang, Y., & Ting, K. C. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass and Bioenergy*, *35*(7), 2455–2469. https://doi.org/10.1016/j.biombioe.2011.02.028

Alawadi, F. (2010). Detection of surface algal blooms using the newly developed algorithm surface algal bloom index (SABI). In C. R. Bostater, Jr., S. P. Mertikas, X. Neyt, & M. Velez-Reyes (Eds.), *Remote Sensing of the Ocean, Sea Ice, and Large Water Regions 2010* (Vol. 7825, p. 782506). https://doi.org/10.1117/12.862096

Anderson, C. R., Berdalet, E., Kudela, R. M., Cusack, C. K., Silke, J., O'Rourke, E., Dugan, D., McCammon, M., Newton, J. A., Moore, S. K., Paige, K., Ruberg, S., Morrison, J. R., Kirkpatrick, B., Hubbard, K., & Morell, J. (2019). Scaling Up From Regional Case Studies to a Global Harmful Algal Bloom Observing System. *Frontiers in Marine Science*, *6*(May). https://doi.org/10.3389/fmars.2019.00250

Ansper, A., & Alikas, K. (2019). Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010064

Appling, A. P., Oliver, S. K., Read, J. S., Sadler, J. M., & Zwart, J. A. (2022). Machine Learning for Understanding Inland Water Quantity, Quality, and Ecology. In *Encyclopedia of Inland Waters* (Vol. 4, pp. 585–606). Elsevier. https://doi.org/10.1016/B978-0-12-819166-8.00121-3

Arar, E. J., & Collins, G. B. (1997). Method 445.0 In Vitro Determination of Chlorophyll a and Pheophytin ain Marine and Freshwater Algae by Fluorescence. In *United States Environmental Protection Agency*

(Issue September). United States Environmental Protection Agency.

Arias-Rodriguez, L. F., Duan, Z., Díaz-Torres, J. de J., Basilio Hazas, M., Huang, J., Kumar, B. U., Tuo, Y., & Disse, M. (2021). Integration of remote sensing and Mexican water quality monitoring system using an extreme learning machine. *Sensors*, *21*(12). https://doi.org/10.3390/s21124118

Auer, M., McDonald, C., Kuczynski, A., Huang, C., & Xue, P. (2021). Management of the Phosphorus–Cladophora Dynamic at a Site on Lake Ontario Using a Multi-Module Bioavailable P Model. *Water*, *13*(3), 375. https://doi.org/10.3390/w13030375

Barreneche, J. M., Guigou, B., Gallego, F., Barbieri, A., Smith, B., Fernández, M., Fernández, V., & Pahlevan, N. (2023). Monitoring Uruguay's freshwaters from space: An assessment of different satellite image processing schemes for chlorophyll-a estimation. *Remote Sensing Applications: Society and Environment*, *29*(January), 100891. https://doi.org/10.1016/j.rsase.2022.100891

Beck, R., Zhan, S., Liu, H., Tong, S., Yang, B., Xu, M., Ye, Z., Huang, Y., Shu, S., Wu, Q., Wang, S., Berling, K., Murray, A., Emery, E., Reif, M., Harwood, J., Young, J., Nietch, C., Macke, D., … Su, H. (2016). Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. *Remote Sensing of Environment*, *178*, 15–30. https://doi.org/10.1016/j.rse.2016.03.002

Binding, C. E., Greenberg, T. A., McCullough, G., Watson, S. B., & Page, E. (2018). An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. *Journal of Great Lakes Research*, *44*(3), 436–446. https://doi.org/10.1016/j.jglr.2018.04.001

Binding, C. E., Greenberg, T. A., Watson, S. B., Rastin, S., & Gould, J. (2015). Long term water clarity changes in North America's Great Lakes from multi-sensor satellite observations. *Limnology and Oceanography*, *60*(6), 1976–1995. https://doi.org/10.1002/lno.10146

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*. https://doi.org/10.1016/j.ecolind.2020.106999

Binding, C. E., Zeng, C., Pizzolato, L., Booth, C., Valipour, R., Fong, P., Zastepa, A., & Pascoe, T. (2023). Reporting on the status, trends, and drivers of algal blooms on Lake of the Woods using satellite-derived bloom indices (2002–2021). *Journal of Great Lakes Research*, *49*(1), 32–43. https://doi.org/10.1016/j.jglr.2022.12.007

Birth, G. S., & McVey, G. R. (1968). Measuring the Color of Growing Turf with a Reflectance

Spectrophotometer 1. *Agronomy Journal*, *60*(6), 640–643. https://doi.org/10.2134/agronj1968.00021962006000060016x

Blagrave, K., Moslenko, L., Khan, U. T., Benoit, N., Howell, T., & Sharma, S. (2022). Heatwaves and storms contribute to degraded water quality conditions in the nearshore of Lake Ontario. *Journal of Great Lakes Research*, *48*(4), 903–913. https://doi.org/10.1016/j.jglr.2022.04.008

Blaustein, J. (1992). The peak near 700 nm on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll. *International Journal of Remote Sensing*, *13*(17), 3367–3373. https://doi.org/10.1080/01431169208904125

Boucher, J., Weathers, K. C., Norouzi, H., & Steele, B. (2018). Assessing the effectiveness of Landsat 8 chlorophyll a retrieval algorithms for regional freshwater monitoring. *Ecological Applications*, *28*(4), 1044–1054. https://doi.org/10.1002/eap.1708

Brivio, P. A., Giardino, C., & Zilioli, E. (2001). Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images. *International Journal of Remote Sensing*, *22*(2–3), 487–502. https://doi.org/10.1080/014311601450059

Buma, W. G., & Lee, S. Il. (2020). Evaluation of Sentinel-2 and Landsat 8 images for estimating Chlorophyll-a concentrations in Lake Chad, Africa. *Remote Sensing*, *12*(15). https://doi.org/10.3390/RS12152437

Candiani, G., Floricioiu, D., Giardino, C., & Rott, H. (2005). Monitoring water quality of the perialpine italian lake garda through multi-temporal meris data. *European Space Agency, (Special Publication) ESA SP*, *597*(597), 401–406.

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., & Xue, K. (2020). A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. *Remote Sensing of Environment*, *248*(June), 111974. https://doi.org/10.1016/j.rse.2020.111974

Cao, Z., Ma, R., Liu, M., Duan, H., Xiao, Q., Xue, K., & Shen, M. (2022). Harmonized Chlorophyll-a Retrievals in Inland Lakes From Landsat-8/9 and Sentinel 2A/B Virtual Constellation Through Machine Learning. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–16. https://doi.org/10.1109/TGRS.2022.3207345

Cao, Z., Ma, R., Melack, J. M., Duan, H., Liu, M., Kutser, T., Xue, K., Shen, M., Qi, T., & Yuan, H. (2022). Landsat observations of chlorophyll-a variations in Lake Taihu from 1984 to 2019. *International Journal of Applied Earth Observation and Geoinformation*, *106*, 102642.

https://doi.org/10.1016/j.jag.2021.102642

Carlson, R. E. (1977). A trophic state index for lakes. *Limnology and Oceanography*, *22*(2), 361–369. https://doi.org/10.4319/lo.1977.22.2.0361

Carmichael, W. W., & Boyer, G. L. (2016). Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful Algae*, *54*, 194–212. https://doi.org/10.1016/j.hal.2016.02.002

Chegoonian, A. M., Pahlevan, N., Zolfaghari, K., Leavitt, P. R., Davies, J. M., Baulch, H. M., & Duguay, C. R. (2023). Comparative Analysis of Empirical and Machine Learning Models for Chla Extraction Using Sentinel-2 and Landsat OLI Data: Opportunities, Limitations, and Challenges. *Canadian Journal of Remote Sensing*, *49*(1). https://doi.org/10.1080/07038992.2023.2215333

Cheng, C., Wei, Y., Lv, G., & Yuan, Z. (2013). Remote estimation of chlorophyll-a concentration in turbid water using a spectral index: a case study in Taihu Lake, China. *Journal of Applied Remote Sensing*, *7*(1), 073465. https://doi.org/10.1117/1.jrs.7.073465

Chomicki, K. M., Taylor, W. D., Brown, C. J. M., Dove, A., Bowen, G. S., & Mohamed, M. N. (2022). Seasonal variation in the influence of environmental drivers on nearshore water quality along an urban northern Lake Ontario shoreline. *Journal of Great Lakes Research*, *48*(4), 914–926. https://doi.org/10.1016/j.jglr.2022.04.011

Concha, J. A., & Schott, J. R. (2016). Retrieval of color producing agents in Case 2 waters using Landsat 8. *Remote Sensing of Environment*, *185*, 95–107. https://doi.org/10.1016/j.rse.2016.03.018

Dall'Olmo, G., & Gitelson, A. A. (2006). Effect of bio-optical parameter variability and uncertainties in reflectance measurements on the remote estimation of chlorophyll-a concentration in turbid productive waters: Modeling results. *Applied Optics*, *45*(15), 3577–3592. https://doi.org/10.1364/AO.45.003577

Dall'Olmo, G., Gitelson, A. A., & Rundquist, D. C. (2003). Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*, *30*(18). https://doi.org/10.1029/2003GL018065

Dekker, A. G., & Peters, S. W. M. (1993). The use of the thematic mapper for the analysis of eutrophic lakes: A case study in the netherlands. *International Journal of Remote Sensing*, *14*(5), 799–821. https://doi.org/10.1080/01431169308904379

Department of the Interior USGS. (2022). Landsat 8-9 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) Collection 2 (C2) Level 2 (L2) Data Format Control Book (DFCB). *Department Inter US Geological Survey*, *2*(September), 1–75.

Deutsch, E. S., Alameddine, I., & El-Fadel, M. (2018). Monitoring water quality in a hypereutrophic reservoir using Landsat ETM+ and OLI sensors: how transferable are the water quality algorithms? *Environmental Monitoring and Assessment*, *190*(3). https://doi.org/10.1007/s10661-018-6506-9

Dierberg, F. E., & Carrlker, N. E. (1994). Field Testing Two Instruments for Remotely Sensing Water Quality in the Tennessee Valley. *Environmental Science and Technology*, *28*(1), 16–25. https://doi.org/10.1021/es00050a004

Doerffer, R. (2015). *Algorithm Theoretical Bases Document (ATBD) for L2 processing of MERIS data of case 2 waters, 4th reprocessing* (Vol. 2).

Dörnhöfer, K., Klinger, P., Heege, T., & Oppelt, N. (2018). Multi-sensor satellite and in situ monitoring of phytoplankton development in a eutrophic-mesotrophic lake. *Science of the Total Environment*, *612*, 1200–1214. https://doi.org/10.1016/j.scitotenv.2017.08.219

Dörnhöfer, K., & Oppelt, N. (2016). Remote sensing for lake research and monitoring - Recent advances. *Ecological Indicators*, *64*, 105–122. https://doi.org/10.1016/j.ecolind.2015.12.009

Dove, A., & Chapra, S. C. (2015). Long-term trends of nutrients and trophic response variables for the Great Lakes. *Limnology and Oceanography*, *60*(2), 696–721. https://doi.org/10.1002/lno.10055

Duan, H., Zhang, Y., Zhang, B., Song, K., Wang, Z., Liu, D., & Li, F. (2008). Estimation of chlorophyll-a concentration and trophic states for inland lakes in Northeast China from Landsat TM data and field spectral measurements. *International Journal of Remote Sensing*, *29*(3), 767–786. https://doi.org/10.1080/01431160701355249

Dube, T., Seaton, D., Shoko, C., & Mbow, C. (2023). Advancements in earth observation for water resources monitoring and management in Africa: A comprehensive review. *Journal of Hydrology*, *623*(May), 129738. https://doi.org/10.1016/j.jhydrol.2023.129738

ECCC, & U.S. EPA. (2022). State of the Great Lakes 2022 Technical Report. In *Cat No. En161-3/1E-PDF. EPA 905-R- 22-004*. binational.net

Estepp, L. R., & Reavie, E. D. (2015). The ecological history of Lake Ontario according to phytoplankton. *Journal of Great Lakes Research*, *41*(3), 669–687. https://doi.org/10.1016/j.jglr.2015.06.005

Feng, L., Dai, Y., Hou, X., Xu, Y., Liu, J., & Zheng, C. (2021). Concerns about phytoplankton bloom trends in global lakes. *Nature*, *590*(7846), E35–E47. https://doi.org/10.1038/s41586-021-03254-3

Feng, L., & Hu, C. (2016). Cloud adjacency effects on top-of-atmosphere radiance and ocean color data products: A statistical assessment. *Remote Sensing of Environment*, *174*, 301–313. https://doi.org/10.1016/j.rse.2015.12.020

Floricioiu, D., Rott, H., Rott, E., Dokulil, M., & Defrancesco, C. (2005). Retrieval of Limnological Parameters of Perialpine Lakes. *Geophysics*, *2004*(April), 4–8.

Gholizadeh, M., Melesse, A., & Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, *16*(8), 1298. https://doi.org/10.3390/s16081298

Gitelson, A. A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., Gurlin, D., & Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, *112*(9), 3582–3593. https://doi.org/10.1016/j.rse.2008.04.015

Gitelson, A. A., Gurlin, D., Moses, W. J., & Barrow, T. (2009). A bio-optical algorithm for the remote estimation of the chlorophyll- a concentration in case 2 waters. *Environmental Research Letters*, *4*(4), 045003. https://doi.org/10.1088/1748-9326/4/4/045003

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gitelson, A. A., & Kondratyev, K. Y. (1991). Optical models of mesotrophic and eutrophic water bodies. *International Journal of Remote Sensing*, *12*(3), 373–385. https://doi.org/10.1080/01431169108929659

Gitelson, A. A., & Yacobi, Y. Z. (1995). Reflectance in the red and near infra-red ranges of the spectrum as tool for remote chlorophyll estimation in inland waters-Lake Kinneret case study. *Eighteenth Convention of Electrical and Electronics Engineers in Israel*, 5.2.6/1-5.2.6/5. https://doi.org/10.1109/EEIS.1995.514184

Gitelson, A., Garbuzov, G., Szilagyi, F., Mittenzwey, K. H., Karnieli, A., & Kaiser, A. (1993). Quantitative remote sensing methods for real-time monitoring of inland waters quality. *International Journal of Remote Sensing*, *14*(7), 1269–1295. https://doi.org/10.1080/01431169308953956

Gitelson, A., Mayo, M., Yacobi, Y. Z., Parparov, A., & Berman, T. (1994). The use of high-spectral-resolution radiometer data for detection of low chlorophyll concentrations in Lake Kinneret. *Journal of Plankton Research*, *16*(8), 993–1002. https://doi.org/10.1093/plankt/16.8.993

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

Gordon, H. R., Brown, O. B., & Jacobs, M. M. (1975). Computed Relationships Between the Inherent and Apparent Optical Properties of a Flat Homogeneous Ocean. *Applied Optics*, *14*(2), 417. https://doi.org/10.1364/ao.14.000417

Gower, J. F. R. R., Doerffer, R., & Borstad, G. A. (1999). Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. *International Journal of Remote Sensing*, *20*(9), 1771–1786. https://doi.org/10.1080/014311699212470

Gower, J., King, S., Borstad, G., & Brown, L. (2005). Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. *International Journal of Remote Sensing*, *26*(9), 2005–2012. https://doi.org/10.1080/01431160500075857

Grendaitė, D., & Stonevičius, E. (2022). Uncertainty of atmospheric correction algorithms for chlorophyll α concentration retrieval in lakes from Sentinel-2 data. *Geocarto International*, *37*(23), 6867–6891. https://doi.org/10.1080/10106049.2021.1958014

Guitelson, A. A., Nikanorov, A. M., Szabo, G. Y., & Szilagyi, F. (1986). Etude de la qualite des eaux de surface par teledetection. *IAHS-AISH Publication*, *157*, 111–121.

Guo, H., Tian, S., Jeanne Huang, J., Zhu, X., Wang, B., & Zhang, Z. (2022). Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. *ISPRS Journal of Photogrammetry and Remote Sensing*, *183*(5), 451–469. https://doi.org/10.1016/j.isprsjprs.2021.11.023

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), 290. https://doi.org/10.3390/ijgi6090290

Hafeez, S., Wong, M. S., Ho, H. C., Nazeer, M., Nichol, J., Abbas, S., Tang, D., Lee, K. H., & Pun, L. (2019). Comparison of machine learning algorithms for retrieval of water quality indicators in case-ii

waters: A case study of hong kong. *Remote Sensing*, *11*(6). https://doi.org/10.3390/rs11060617

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

He, Y., Jin, S., & Shang, W. (2021). Water quality variability and related factors along the yangtze river using landsat-8. *Remote Sensing*, *13*(12), 1–20. https://doi.org/10.3390/rs13122241

Hewson, R. D., Cudahy, T. J., & Huntington, J. F. (2001). Geologic and alteration mapping at Mt Fitton, South Australia, using ASTER satellite-borne data. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, *2*, 724–726. https://doi.org/10.1109/IGARSS.2001.976615

Higgins, S. N., Pennuto, C. M., Howell, E. T., Lewis, T. W., & Makarewicz, J. C. (2012). Urban influences on Cladophora blooms in Lake Ontario. *Journal of Great Lakes Research*, *38*(SUPPL.4), 116–123. https://doi.org/10.1016/j.jglr.2011.11.017

Hiriart-Baer, V. P., Milne, J., & Charlton, M. N. (2009a). Water quality trends in Hamilton Harbour: Two decades of change in nutrients and chlorophyll a. *Journal of Great Lakes Research*, *35*(2), 293–301. https://doi.org/10.1016/j.jglr.2008.12.007

Hiriart-Baer, V. P., Milne, J., & Charlton, M. N. (2009b). Water quality trends in Hamilton Harbour: Two decades of change in nutrients and chlorophyll a. *Journal of Great Lakes Research*, *35*(2), 293–301. https://doi.org/10.1016/j.jglr.2008.12.007

Ho, J. C., Michalak, A. M., & Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature*, *574*(7780), 667–670. https://doi.org/10.1038/s41586-019-1648-7

Howell, E. (2018). Influences on Water Quality and Abundance of Cladophora, a Shore-Fouling Green Algae, over Urban Shoreline in Lake Ontario. *Water*, *10*(11), 1569. https://doi.org/10.3390/w10111569

Howell, E. T., & Benoit, N. (2021a). Loading and lake circulation structures recurrent patterns of water quality on the Toronto – Mississauga waterfront of Lake Ontario. *Journal of Great Lakes Research*, *47*(2), 323–342. https://doi.org/10.1016/j.jglr.2020.10.012

Howell, E. T., & Benoit, N. (2021b). Nutrient footprints on the Toronto-Mississauga waterfront of Lake

Ontario. *Journal of Great Lakes Research*, *47*(2), 343–365. https://doi.org/10.1016/j.jglr.2020.11.009

Hu, C., Feng, L., Lee, Z., Franz, B. A., Bailey, S. W., Werdell, P. J., & Proctor, C. W. (2019). Improving Satellite Global Chlorophyll a Data Products Through Algorithm Refinement and Data Recovery. *Journal of Geophysical Research: Oceans*, *124*(3), 1524–1543. https://doi.org/10.1029/2019JC014941

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans*, *117*(1). https://doi.org/10.1029/2011JC007395

Hu, M., Ma, R., Cao, Z., Xiong, J., & Xue, K. (2021). Remote estimation of trophic state index for inland waters using landsat-8 oli imagery. *Remote Sensing*, *13*(10), 1988. https://doi.org/10.3390/rs13101988

Huang, A., Rao, Y. R., & Zhang, W. (2012). On Recent Trends in Atmospheric and Limnological Variables in Lake Ontario. *Journal of Climate*, *25*(17), 5807–5816. https://doi.org/10.1175/JCLI-D-11-00495.1

Hui, Y., Zhu, Z., Atkinson, J. F., & Saharia, A. M. (2021). Impacts of phosphorus loading temporal pattern on benthic algae growth in Lake Ontario. *Journal of Hydrology*, *598*(January), 126449. https://doi.org/10.1016/j.jhydrol.2021.126449

Ilteralp, M., Ariman, S., & Aptoula, E. (2022). A deep multitask semisupervised learning approach for chlorophyll-a retrieval from remote sensing images. *Remote Sensing*, *14*(1), 1–13. https://doi.org/10.3390/rs14010018

Jabbari, A., Valipour, R., Ackerman, J. D., & Rao, Y. R. (2023). Nearshore-offshore exchanges by enhanced turbulent mixing along the north shore of Lake Ontario. *Journal of Great Lakes Research*, *49*(3), 596–607. https://doi.org/10.1016/j.jglr.2023.03.010

Kahru, M., & Elmgren, R. (2014). Multidecadal time series of satellite-detected accumulations of cyanobacteria in the Baltic Sea. *Biogeosciences*, *11*(13), 3619–3633. https://doi.org/10.5194/bg-11-3619-2014

Kahru, M., & Mitchell, B. G. (1998). Spectral reflectance and absorption of a massive red tide off southern California. *Journal of Geophysical Research: Oceans*, *103*(C10), 21601–21609. https://doi.org/10.1029/98JC01945

Khan, R. M., Salehi, B., Mahdianpari, M., Mohammadimanesh, F., Mountrakis, G., & Quackenbush, L. J.

(2021). A meta-analysis on harmful algal bloom (Hab) detection and monitoring: A remote sensing perspective. *Remote Sensing*, *13*(21), 1–41. https://doi.org/10.3390/rs13214347

Koponen, S., Attila, J., Pulliainen, J., Kallio, K., Pyhälahti, T., Lindfors, A., Rasmus, K., & Hallikainen, M. (2007). A case study of airborne and satellite remote sensing of a spring bloom event in the Gulf of Finland. *Continental Shelf Research*, *27*(2), 228–244. https://doi.org/10.1016/j.csr.2006.10.006

Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., & Casal, G. (2016). Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. *Remote Sensing*, *8*(6), 497. https://doi.org/10.3390/rs8060497

Le, C., Hu, C., Cannizzaro, J., English, D., Muller-Karger, F., & Lee, Z. (2013). Evaluation of chlorophyll-a remote sensing algorithms for an optically complex estuary. *Remote Sensing of Environment*, *129*, 75–89. https://doi.org/10.1016/j.rse.2012.11.001

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). https://doi.org/10.3390/hydrology10050110

Lesht, B. M., Barbiero, R. P., & Warren, G. J. (2013). A band-ratio algorithm for retrieving open-lake chlorophyll values from satellite observations of the Great Lakes. *Journal of Great Lakes Research*, *39*(1), 138–152. https://doi.org/10.1016/j.jglr.2012.12.007

Lesht, B. M., Barbiero, R. P., & Warren, G. J. (2016). Verification of a simple band ratio algorithm for retrieving Great Lakes open water surface chlorophyll concentrations from satellite observations. *Journal of Great Lakes Research*, *42*(2), 448–454. https://doi.org/10.1016/j.jglr.2015.12.013

Li, Y., Zhang, Y., Shi, K., Zhou, Y., Zhang, Y., Liu, X., & Guo, Y. (2018). Spatiotemporal dynamics of chlorophyll-a in a large reservoir as derived from Landsat 8 OLI data: understanding its driving and restrictive factors. *Environmental Science and Pollution Research*, *25*(2), 1359–1374. https://doi.org/10.1007/s11356-017-0536-7

Li, Y., Zhou, Q., Zhang, Y., Li, J., & Shi, K. (2021). Research Trends in the Remote Sensing of Phytoplankton Blooms: Results from Bibliometrics. *Remote Sensing*, *13*(21), 4414. https://doi.org/10.3390/rs13214414

Lillesand, T. M., Johnson, W. L., & Deuell, R. L. (1983). Use of landsat data to predict the trophic state of Minnesota lakes. *Photogrammetric Engineering and Remote Sensing*, *49*(2), 219–229.

Lisboa, F., Brotas, V., Santos, F. D., Kuikka, S., Kaikkonen, L., & Maeda, E. E. (2020). Spatial variability and detection levels for chlorophyll-a estimates in high latitude lakes using landsat imagery. *Remote Sensing*, *12*(18). https://doi.org/10.3390/RS12182898

Lobo, F. de L., Nagel, G. W., Maciel, D. A., de Carvalho, L. A. S., Martins, V. S., Barbosa, C. C. F., & de Moraes Novo, E. M. L. (2021). Algaemap: Algae bloom monitoring application for inland waters in Latin America. *Remote Sensing*, *13*(15). https://doi.org/10.3390/rs13152874

Ma, J., He, F., Qi, T., Sun, Z., Shen, M., Cao, Z., Meng, D., Duan, H., & Luo, J. (2022). Thirty-Four-Year Record (1987–2021) of the Spatiotemporal Dynamics of Algal Blooms in Lake Dianchi from Multi-Source Remote Sensing Insights. *Remote Sensing*, *14*(16). https://doi.org/10.3390/rs14164000

Maciel, F. P., Haakonsson, S., Ponce de León, L., Bonilla, S., & Pedocchi, F. (2023). Challenges for chlorophyll-a remote sensing in a highly variable turbidity estuary, an implementation with sentinel-2. *Geocarto International*, *38*(1). https://doi.org/10.1080/10106049.2022.2160017

Maeda, E. E., Lisboa, F., Kaikkonen, L., Kallio, K., Koponen, S., Brotas, V., & Kuikka, S. (2019). Temporal patterns of phytoplankton phenology across high latitude lakes unveiled by long-term time series of satellite data. *Remote Sensing of Environment*, *221*(June 2018), 609–620. https://doi.org/10.1016/j.rse.2018.12.006

Malkin, S. Y., Dove, A., Depew, D., Smith, R. E., Guildford, S. J., & Hecky, R. E. (2010). Spatiotemporal patterns of water quality in Lake Ontario and their implications for nuisance growth of Cladophora. *Journal of Great Lakes Research*, *36*(3), 477–489. https://doi.org/10.1016/j.jglr.2010.06.007

Mamun, M., Ferdous, J., & An, K.-G. (2021). Empirical Estimation of Nutrient, Organic Matter and Algal Chlorophyll in a Drinking Water Reservoir Using Landsat 5 TM Data. *Remote Sensing*, *13*(12), 2256. https://doi.org/10.3390/rs13122256

Mansaray, A. S., Dzialowski, A. R., Martin, M. E., Wagner, K. L., Gholizadeh, H., & Stoodley, S. H. (2021). Comparing planetscope to landsat-8 and sentinel-2 for sensing water quality in reservoirs in agricultural watersheds. *Remote Sensing*, *13*(9), 1–20. https://doi.org/10.3390/rs13091847

Markogianni, V., Kalvas, D., Petropoulos, G. P., & Dimitriou, E. (2018). An appraisal of the potential of Landsat 8 in estimating chlorophyll-a, ammonium concentrations and other water quality indicators. *Remote Sensing*, *10*(7), 1–22. https://doi.org/10.3390/rs10071018

Markovic, S., Liang, A., Watson, S. B., Depew, D., Zastepa, A., Surana, P., Byllaardt, J. Vanden, Arhonditsis, G., & Dittrich, M. (2019). Reduction of industrial iron pollution promotes phosphorus

internal loading in eutrophic Hamilton Harbour, Lake Ontario, Canada. *Environmental Pollution*, *252*, 697–705. https://doi.org/10.1016/j.envpol.2019.05.124

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., 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

Matus-Hernández, M. Á., Hernández-Saavedra, N. Y., & Octavio Martínez-Rincón, R. (2018). Predictive performance of regression models to estimate Chlorophyll - A concentration based on Landsat imagery. *PLoS ONE*, *13*(10), 1–18. https://doi.org/10.1371/journal.pone.0205682

Mayo, M., Gitelson, A., Yacobi, Y. Z., & Ben-Avraham, Z. (1995). Chlorophyll distribution in Lake Kinneret determined from Landsat Thematic Mapper data. *International Journal of Remote Sensing*, *16*(1), 175–182. https://doi.org/10.1080/01431169508954386

MECP. (2015). *The determination of chlorophyll in river and lake samples by spectrophotometry (RCHLO-E3169)*. *6*, 1–37.

MECP. (2016). The determination of chlorophylls A and B and total chlorophyll A in river and lake samples by diode array detector-liquid chromatography-tandem mass spectrometry (E3508). *Laboratory Services Branch*, *3508*.

MECP. (2023). *The Determination Of Chlorophylls A And B And Total Chlorophyll A In River And Lake Samples By Diode Array Detector – Liquid Chromatography – Tandem Mass Spectrometry (E3508)*.

Mehta, A., & Flores, A. (2018). *Processing Satellite Imagery for Monitoring Water Quality*. NASA Applied Remote Sensing Training Program (ARSET). https://appliedsciences.nasa.gov/join-mission/training/english/arset-processing-satellite-imagery-monitoring-water-quality

Mishra, D. R., Ogashawara, I., & Gitelson, A. A. (2017). Bio-optical Modeling and Remote Sensing of Inland Waters. In *Bio-optical Modeling and Remote Sensing of Inland Waters*.

Mishra, S., & Mishra, D. R. (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of*

Environment, *117*, 394–406. https://doi.org/10.1016/j.rse.2011.10.016

Mishra, S. P., & Jena, J. (2012). Effects of Variable Inflow from Northern Major Rivers into the Chilika Lagoon , Odisha , India. *International Journal of Lakes and Rivers*, *5*(2), 123–132.

Mittenzwey, K. -H, Ullrich, S., Gitelson, A. A., & Kondratiev, K. Y. (1992). Determination of chlorophyll a of inland waters on the basis of spectral reflectance. *Limnology and Oceanography*, *37*(1), 147–149. https://doi.org/10.4319/lo.1992.37.1.0147

Mohamed, M. N., Wellen, C., Parsons, C. T., Taylor, W. D., Arhonditsis, G., Chomicki, K. M., Boyd, D., Weidman, P., Mundle, S. O. C., Van Cappellen, P., Sharpley, A. N., & Haffner, D. G. (2019). Understanding and managing the re-eutrophication of lake erie: Knowledge gaps and research priorities. *Freshwater Science*, *38*(4), 675–691. https://doi.org/10.1086/705915

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

Mortula, M., Ali, T., Bachir, A., Elaksher, A., & Abouleish, M. (2020). Towards Monitoring of Nutrient Pollution in Coastal Lake Using Remote Sensing and Regression Analysis. *Water*, *12*(7), 1954. https://doi.org/10.3390/w12071954

Moses, W. J., Gitelson, A. A., Berdnikov, S., & Povazhnyy, V. (2009). Satellite estimation of chlorophyll-a concentration using the red and NIR bands of MERISThe azov sea case study. *IEEE Geoscience and Remote Sensing Letters*, *6*(4), 845–849. https://doi.org/10.1109/LGRS.2009.2026657

Moses, W. J., Sterckx, S., Montes, M. J., De Keukelaere, L., & Knaeps, E. (2017). Atmospheric Correction for Inland Waters. In *Bio-optical Modeling and Remote Sensing of Inland Waters* (pp. 69–100). Elsevier. https://doi.org/10.1016/B978-0-12-804644-9.00003-3

Munawar, M., & Fitzpatrick, M. A. J. (2018). Eutrophication in three Canadian Areas of Concern: Phytoplankton and major nutrient interactions. *Aquatic Ecosystem Health and Management*, *21*(4), 421–437. https://doi.org/10.1080/14634988.2018.1530895

Munawar, M., Fitzpatrick, M., Niblock, H., Kling, H., Rozon, R., & Lorimer, J. (2017). Phytoplankton ecology of a culturally eutrophic embayment: Hamilton Harbour, Lake Ontario. *Aquatic Ecosystem Health and Management*, *20*(3), 201–213. https://doi.org/10.1080/14634988.2017.1307678

National Laboratory for Environmental Testing, B. (2021). *Standard Operating Procedure for the Analysis of Chlorophyll a in Natural Waters by Spectrophotometric Determination (Sop B0258w)*. ECCC.

Nazeer, M., & Nichol, J. E. (2016). Development and application of a remote sensing-based Chlorophyll-a concentration prediction model for complex coastal waters of Hong Kong. *Journal of Hydrology*, *532*, 80–89. https://doi.org/10.1016/j.jhydrol.2015.11.037

Neville, R. A., & Gower, J. F. R. (1977). Passive remote sensing of phytoplankton via chlorophyll α fluorescence. *Journal of Geophysical Research*, *82*(24), 3487–3493. https://doi.org/10.1029/JC082i024p03487

Niroumand-Jadidi, M., Bovolo, F., Bresciani, M., Gege, P., & Giardino, C. (2022). Water Quality Retrieval from Landsat-9 (OLI-2) Imagery and Comparison to Sentinel-2. *Remote Sensing*, *14*(18), 4596. https://doi.org/10.3390/rs14184596

NOAA. (2024). *Lake Ontario | National Marine Ecosystem Status*. https://ecowatch.noaa.gov/index.php/regions/great-lakes/lake-ontario

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: Oceans*, *103*(C11), 24937–24953. https://doi.org/10.1029/98JC02160

O'Reilly, J. E., & Werdell, P. J. (2019). Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Remote Sensing of Environment*, *229*(October 2018), 32–47. https://doi.org/10.1016/j.rse.2019.04.021

Ogashawara, I., Kiel, C., Jechow, A., Kohnert, K., Ruhtz, T., Grossart, H. P., Hölker, F., Nejstgaard, J. C., Berger, S. A., & Wollrab, S. (2021). The use of sentinel-2 for chlorophyll-A spatial dynamics assessment: A comparative study on different lakes in northern Germany. *Remote Sensing*, *13*(8), 1–26. https://doi.org/10.3390/rs13081542

Olivetti, D., Roig, H., Martinez, J. M., Borges, H., Ferreira, A., Casari, R., Salles, L., & Malta, E. (2020). Low-cost unmanned aerial multispectral imagery for siltation monitoring in reservoirs. *Remote Sensing*, *12*(11). https://doi.org/10.3390/rs12111855

Pahlevan, N., Mangin, A., Balasubramanian, S. V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikaza, J., Kratzer, S., Lehmann, M. K., Ligi, M., … Warren, M. (2021). ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. *Remote Sensing of Environment*, *258*(August 2020), 112366. https://doi.org/10.1016/j.rse.2021.112366

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M. K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., … Ruiz-Verdù, A. (2022). Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sensing of Environment*, *270*. https://doi.org/10.1016/j.rse.2021.112860

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N., & Stumpf, R. (2020a). 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

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N., & Stumpf, R. (2020b). 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*(December 2019), 111604. https://doi.org/10.1016/j.rse.2019.111604

Paltsev, A., & Creed, I. F. (2022). Are Northern Lakes in Relatively Intact Temperate Forests Showing Signs of Increasing Phytoplankton Biomass? *Ecosystems*, *25*(3), 727–755. https://doi.org/10.1007/s10021-021-00684-y

Pereira-Sandoval, M., Ruescas, A., Urrego, P., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J., & Moreno, J. (2019). Evaluation of atmospheric correction algorithms over spanish inland waters for sentinel-2 multi spectral imagery data. *Remote Sensing*, *11*(12), 1–23. https://doi.org/10.3390/rs11121469

Philipson, P., Kratzer, S., Ben Mustapha, S., Strömbeck, N., & Stelzer, K. (2016). Satellite-based water quality monitoring in Lake Vänern, Sweden. *International Journal of Remote Sensing*, *37*(16), 3938–3960. https://doi.org/10.1080/01431161.2016.1204480

Pirasteh, S., Mollaee, S., Fatholahi, S. N., & Li, J. (2020). Estimation of Phytoplankton Chlorophyll-a Concentrations in the Western Basin of Lake Erie Using Sentinel-2 and Sentinel-3 Data. *Canadian Journal of Remote Sensing*, *46*(5), 585–602. https://doi.org/10.1080/07038992.2020.1823825

Poddar, S., Chacko, N., & Swain, D. (2019). Estimation of Chlorophyll-a in Northern Coastal Bay of Bengal Using Landsat-8 OLI and Sentinel-2 MSI Sensors. *Frontiers in Marine Science*, *6*.

https://doi.org/10.3389/fmars.2019.00598

Pokrzywinski, K., Johansen, R., Reif, M., Bourne, S., Hammond, S., & Fernando, B. (2022). Remote sensing of the cyanobacteria life cycle: A mesocosm temporal assessment of a Microcystis sp. bloom using coincident unmanned aircraft system (UAS) hyperspectral imagery and ground sampling efforts. *Harmful Algae*, *117*(May), 102268. https://doi.org/10.1016/j.hal.2022.102268

Rodríguez-López, L., Duran-Llacer, I., González-Rodríguez, L., Abarca-del-Rio, R., Cárdenas, R., Parra, O., Martínez-Retureta, R., & Urrutia, R. (2020). Spectral analysis using LANDSAT images to monitor the chlorophyll-a concentration in Lake Laja in Chile. *Ecological Informatics*, *60*(June). https://doi.org/10.1016/j.ecoinf.2020.101183

Rouse, J. W., Haas, R. H., Deering, D. W., Schell, J. A., & Harlan, J. C. (1974). Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation - NASA Technical Reports Server (NTRS). In *Earth Resources and Remote Sensing*.

Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. *Nasa ERTS Symposium*, *351*(1), 309–313.

Saberioon, M., Brom, J., Nedbal, V., Souček, P., & Císař, P. (2020). Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. *Ecological Indicators*, *113*(March), 106236. https://doi.org/10.1016/j.ecolind.2020.106236

Sagan, V., Peterson, K. T., Maimaitijiang, M., Sidike, P., Sloan, J., Greeling, B. A., Maalouf, S., & 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*(August 2019), 103187. https://doi.org/10.1016/j.earscirev.2020.103187

Salem, S., Higa, H., Kim, H., Kobayashi, H., Oki, K., & Oki, T. (2017). Assessment of Chlorophyll-a Algorithms Considering Different Trophic Statuses and Optimal Bands. *Sensors*, *17*(8), 1746. https://doi.org/10.3390/s17081746

Saranathan, A. M., Smith, B., & Pahlevan, N. (2023). Per-Pixel Uncertainty Quantification and Reporting for Satellite-Derived Chlorophyll-a Estimates via Mixture Density Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–18. https://doi.org/10.1109/TGRS.2023.3234465

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

Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., & Werdell, P. J. (2018). Performance metrics for the assessment of satellite data products: an ocean color case study. *Optics Express*, *26*(6), 7404. https://doi.org/10.1364/oe.26.007404

Seegers, B. N., Werdell, P. J., Vandermeulen, R. A., Salls, W., Stumpf, R. P., Schaeffer, B. A., Owens, T. J., Bailey, S. W., Scott, J. P., & Loftin, K. A. (2021). Satellites for long-term monitoring of inland U.S. lakes: The MERIS time series and application for chlorophyll-a. *Remote Sensing of Environment*, *266*(August), 112685. https://doi.org/10.1016/j.rse.2021.112685

Sherman, J., Tzortziou, M., Turner, K. J., Goes, J., & Grunert, B. (2023). Chlorophyll dynamics from Sentinel-3 using an optimized algorithm for enhanced ecological monitoring in complex urban estuarine waters. *International Journal of Applied Earth Observation and Geoinformation*, *118*(December 2022), 103223. https://doi.org/10.1016/j.jag.2023.103223

Shi, K., Zhang, Y., Qin, B., & Zhou, B. (2019). Remote sensing of cyanobacterial blooms in inland waters: present knowledge and future challenges. *Science Bulletin*, *64*(20), 1540–1556. https://doi.org/10.1016/j.scib.2019.07.002

Shi, K., Zhang, Y., Song, K., Liu, M., Zhou, Y., Zhang, Y., Li, Y., Zhu, G., & Qin, B. (2019). A semi-analytical approach for remote sensing of trophic state in inland waters: Bio-optical mechanism and application. *Remote Sensing of Environment*, *232*, 111349. https://doi.org/10.1016/j.rse.2019.111349

Smith, M. E., Robertson Lain, L., & Bernard, S. (2018). An optimized Chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. *Remote Sensing of Environment*, *215*(January), 217–227. https://doi.org/10.1016/j.rse.2018.06.002

Soomets, T., Uudeberg, K., Jakovels, D., Brauns, A., Zagars, M., & Kutser, T. (2020). Validation and Comparison of Water Quality Products in Baltic Lakes Using Sentinel-2 MSI and Sentinel-3 OLCI Data. *Sensors*, *20*(3), 742. https://doi.org/10.3390/s20030742

Soontiens, N., Binding, C., Fortin, V., Mackay, M., & Rao, Y. R. (2019). Algal bloom transport in Lake Erie using remote sensing and hydrodynamic modelling: Sensitivity to buoyancy velocity and initial vertical distribution. *Journal of Great Lakes Research*, *45*(3), 556–572. https://doi.org/10.1016/j.jglr.2018.10.003

Sòria-Perpinyà, X., Delegido, J., Urrego, E. P., Ruíz-Verdú, A., Soria, J. M., Vicente, E., & Moreno, J. (2022). Assessment of Sentinel-2-MSI Atmospheric Correction Processors and In Situ Spectrometry Waters Quality Algorithms. *Remote Sensing*, *14*(19), 4794. https://doi.org/10.3390/rs14194794

Sòria-Perpinyà, X., Vicente, E., Urrego, P., Pereira-Sandoval, M., Tenjo, C., Ruíz-Verdú, A., Delegido, J., Soria, J. M., Peña, R., & Moreno, J. (2021). Validation of Water Quality Monitoring Algorithms for Sentinel-2 and Sentinel-3 in Mediterranean Inland Waters with In Situ Reflectance Data. *Water*, *13*(5), 686. https://doi.org/10.3390/w13050686

Soriano-González, J., Angelats, E., Fernández-Tejedor, M., Diogene, J., & Alcaraz, C. (2019). First results of phytoplankton spatial dynamics in two NW-Mediterranean bays from chlorophyll-A estimates using Sentinel 2: Potential implications for aquaculture. *Remote Sensing*, *11*(15). https://doi.org/10.3390/rs11151756

Souza, A. P. D., Teodoro, P. E., Teodoro, L. P. R., Taveira, A. C., de Oliveira-Júnior, J. F., Della-Silva, J. L., Baio, F. H. R., Lima, M., & da Silva Junior, C. A. (2021). Application of remote sensing in environmental impact assessment: a case study of dam rupture in Brumadinho, Minas Gerais, Brazil. *Environmental Monitoring and Assessment*, *193*(9). https://doi.org/10.1007/s10661-021-09417-z

Strickland, J. D. ., & Parsons, T. R. (1972). A practical handbook of seawater analysis. Second Edition, Bulletin 167. *Fish. Res. Board Canada, Ottawa.*

Stumpf, R. P., Davis, T. W., Wynne, T. T., Graham, J. L., Loftin, K. A., Johengen, T. H., Gossiaux, D., Palladino, D., & Burtner, A. (2016). Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. *Harmful Algae*, *54*, 160–173. https://doi.org/10.1016/j.hal.2016.01.005

Sun, D., Hu, C., Qiu, Z., & Shi, K. (2015). Estimating phycocyanin pigment concentration in productive inland waters using Landsat measurements: A case study in Lake Dianchi. *Optics Express*, *23*(3), 3055. https://doi.org/10.1364/oe.23.003055

Tan, W., Liu, P., Liu, Y., Yang, S., & Feng, S. (2017). A 30-Year Assessment of Phytoplankton Blooms in Erhai Lake Using Landsat Imagery: 1987 to 2016. *Remote Sensing*, *9*(12), 1265. https://doi.org/10.3390/rs9121265

Tavares, M. H., Lins, R. C., Harmel, T., Fragoso Jr., C. R., Martínez, J.-M., & Motta-Marques, D. (2021). Atmospheric and sunglint correction for retrieving chlorophyll-a in a productive tropical estuarine-lagoon system using Sentinel-2 MSI imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *174*(January), 215–236. https://doi.org/10.1016/j.isprsjprs.2021.01.021

Tavora, J., Jiang, B., Kiffney, T., Bourdin, G., Gray, P. C., Carvalho, L. S., Hesketh, G., Schild, K. M., Souza, L. F., Brady, D. C., & Boss, E. (2023). Recipes for the Derivation of Water Quality Parameters Using the High-Spatial-Resolution Data from Sensors on Board Sentinel-2A, Sentinel-2B, Landsat-

5, Landsat-7, Landsat-8, and Landsat-9 Satellites. *Journal of Remote Sensing*, *3*(Table 1), 1–19. https://doi.org/10.34133/remotesensing.0049

Tian, S., Guo, H., Huang, J. J., Zhu, X., & Zhang, Z. (2022). Comprehensive comparison performances of Landsat-8 atmospheric correction methods for inland and coastal waters. *Geocarto International*, *37*(27), 15302–15323. https://doi.org/10.1080/10106049.2022.2097320

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First experiences in mapping lakewater quality parameters with sentinel-2 MSI imagery. *Remote Sensing*, *8*(8), 640. https://doi.org/10.3390/rs8080640

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Tuygun, G. T., Salgut, S., & Elçi, A. (2023). Long-term spatial-temporal monitoring of eutrophication in Lake Burdur using remote sensing data. *Water Science and Technology*, *87*(9), 2184–2194. https://doi.org/10.2166/wst.2023.113

U.S. EPA, & Government of Canada. (2022). *State of the Great Lakes 2022 Report: An overview of the status and trends of the Great Lakes ecosystem*. 40. https://binational.net/2022/07/29/sogl-edgl-2022/

Vélez-Nicolás, M., García-López, S., Barbero, L., Ruiz-Ortiz, V., & Sánchez-Bellón, Á. (2021). Applications of unmanned aerial systems (UASs) in hydrology: A review. *Remote Sensing*, *13*(7). https://doi.org/10.3390/rs13071359

Villota-González, F. H., Sulbarán-Rangel, B., Zurita-Martínez, F., Gurubel-Tun, K. J., & Zúñiga-Grajeda, V. (2023). Assessment of Machine Learning Models for Remote Sensing of Water Quality in Lakes Cajititlán and Zapotlán, Jalisco – Mexico. *Remote Sensing*, *15*(23), 5505. https://doi.org/10.3390/rs15235505

Wang, D., Tang, B.-H., Fu, Z., Huang, L., Li, M., Chen, G., & Pan, X. (2022). Estimation of Chlorophyll-A Concentration with Remotely Sensed Data for the Nine Plateau Lakes in Yunnan Province. *Remote Sensing*, *14*(19), 4950. https://doi.org/10.3390/rs14194950

Wang, Z., Li, B., & Li, L. (2019). Research on Water Quality Detection Technology Based on Multispectral Remote Sensing. *IOP Conference Series: Earth and Environmental Science*, *237*(3), 032087. https://doi.org/10.1088/1755-1315/237/3/032087

Warren, M. A., Simis, S. G. H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E.,

Giardino, C., & Ansper, A. (2019). Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sensing of Environment*, *225*(March), 267–289. https://doi.org/10.1016/j.rse.2019.03.018

Warren, M. A., Simis, S. G. H., & Selmes, N. (2021). Complementary water quality observations from high and medium resolution Sentinel sensors by aligning chlorophyll-a and turbidity algorithms. *Remote Sensing of Environment*, *265*(April). https://doi.org/10.1016/j.rse.2021.112651

Xiao, Y., Chen, J., Xu, Y., Guo, S., Nie, X., Guo, Y., Li, X., Hao, F., & Fu, Y. H. (2023). Monitoring of chlorophyll-a and suspended sediment concentrations in optically complex inland rivers using multisource remote sensing measurements. *Ecological Indicators*, *155*(October), 111041. https://doi.org/10.1016/j.ecolind.2023.111041

Yan, K., Li, J., Zhao, H., Wang, C., Hong, D., Du, Y., Mu, Y., Tian, B., Xie, Y., Yin, Z., Zhang, F., & Wang, S. (2022). Deep Learning-Based Automatic Extraction of Cyanobacterial Blooms from Sentinel-2 MSI Satellite Data. *Remote Sensing*, *14*(19). https://doi.org/10.3390/rs14194763

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, W., Matsushita, B., Chen, J., Fukushima, T., & Ma, R. (2010). An enhanced three-band index for estimating chlorophyll-a in turbid case-II waters: Case studies of Lake Kasumigaura, Japan, and Lake Dianchi, China. *IEEE Geoscience and Remote Sensing Letters*, *7*(4), 655–659. https://doi.org/10.1109/LGRS.2010.2044364

Yang, X., Jiang, Y., Deng, X., Zheng, Y., & Yue, Z. (2020). Temporal and Spatial Variations of Chlorophyll a Concentration and Eutrophication Assessment (1987–2018) of Donghu Lake in Wuhan Using Landsat Images. *Water*, *12*(8), 2192. https://doi.org/10.3390/w12082192

Yao, H., Huang, Y., Wei, Y., Zhong, W., & Wen, K. (2021). Retrieval of Chlorophyll-a Concentrations in the Coastal Waters of the Beibu Gulf in Guangxi Using a Gradient-Boosting Decision Tree Model. *Applied Sciences*, *11*(17), 7855. https://doi.org/10.3390/app11177855

Yasuoka, Y., & Miyazaki, T. (1982). Remote Sensing of Water Quality in the Lake. *Journal of the Remote Sensing Society of Japan*, *2*(3), 51–62.

Yin, Z., Li, J., Zhang, B., Liu, Y., Yan, K., Gao, M., Xie, Y., Zhang, F., & Wang, S. (2023). Increase in chlorophyll-a concentration in Lake Taihu from 1984 to 2021 based on Landsat observations. *Science*

of the Total Environment, *873*(February), 162168. https://doi.org/10.1016/j.scitotenv.2023.162168

Yuan, X., Wang, S., Fan, F., Dong, Y., Li, Y., Lin, W., & Zhou, C. (2022). Spatiotemporal dynamics and anthropologically dominated drivers of chlorophyll-a, TN and TP concentrations in the Pearl River Estuary based on retrieval algorithm and random forest regression. *Environmental Research*, *215*(August). https://doi.org/10.1016/j.envres.2022.114380

Zarco-Tejada, P. J., & Ustin, S. L. (2001). Modeling canopy water content for carbon estimates from MODIS data at land EOS validation sites. *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, *1*, 342–344. https://doi.org/10.1109/IGARSS.2001.976152

Zastepa, A., & Chemali, C. (2021). Bloom announcement: Late season cyanobacterial blooms co-dominated by Microcystis flos-aquae, Lyngbya birgei, and Aphanizomenon flos-aquae complex in Hamilton Harbour (Lake Ontario), an area of concern impacted by industrial effluent and residential waste. *Data in Brief*, *35*. https://doi.org/10.1016/j.dib.2021.106800

Zeng, C., & Binding, C. E. (2021). Consistent multi-mission measures of inland water algal bloom spatial extent using meris, modis and olci. *Remote Sensing*, *13*(17), 3349. https://doi.org/10.3390/rs13173349

Zhang, F., Li, J., Yan, B., Yu, J., Wang, C., Wang, S., Shen, Q., Wu, Y., & Zhang, B. (2021). Tracking historical chlorophyll-a change in the guanting reservoir, Northern China, based on landsat series inter-sensor normalization. *International Journal of Remote Sensing*, *42*(10), 3918–3937. https://doi.org/10.1080/01431161.2021.1875149

Zhang, H., Xue, B., Wang, G., Zhang, X., & Zhang, Q. (2022). Deep Learning-Based Water Quality Retrieval in an Impounded Lake Using Landsat 8 Imagery: An Application in Dongping Lake. *Remote Sensing*, *14*(18). https://doi.org/10.3390/rs14184505

Zolfaghari, K., Pahlevan, N., Simis, S. G. H., Shea, R. E. O., & Duguay, C. R. (2022). Sensitivity of remotely sensed pigment concentration via Mixture Density Networks (MDNs) to uncertainties from atmospheric correction. *Journal of Great Lakes Research*, *xxxx*. https://doi.org/10.1016/j.jglr.2022.12.010

Appendices

Appendix I

Supplementary Materials from Chapter 3

This section provides supplementary visualizations related to the data and results. The scatter plot in Figure A1.1 illustrates the temporal distribution of in-situ Chl-*a* matchups, colour-coded based on corresponding satellite imagery. Figure A1.2 displays a heatmap emphasizing the significance of each scheme in each data category with respect to in-situ Chl-*a* concentration, with colour variations representing different R^2 performances. Four tables are also presented in this section. Table A1.1 summarizes various definitions of the Carlson's Trophic State Index (TSI) that exist in the literature. Table A1.2 offers quantitative details of the in-situ data utilized, while Table A1.3 describes each of the atmospheric correction processors employed. Table A1.4 delivers a comprehensive description of the RS-derived Chl-a models developed, along with their performance metrics.

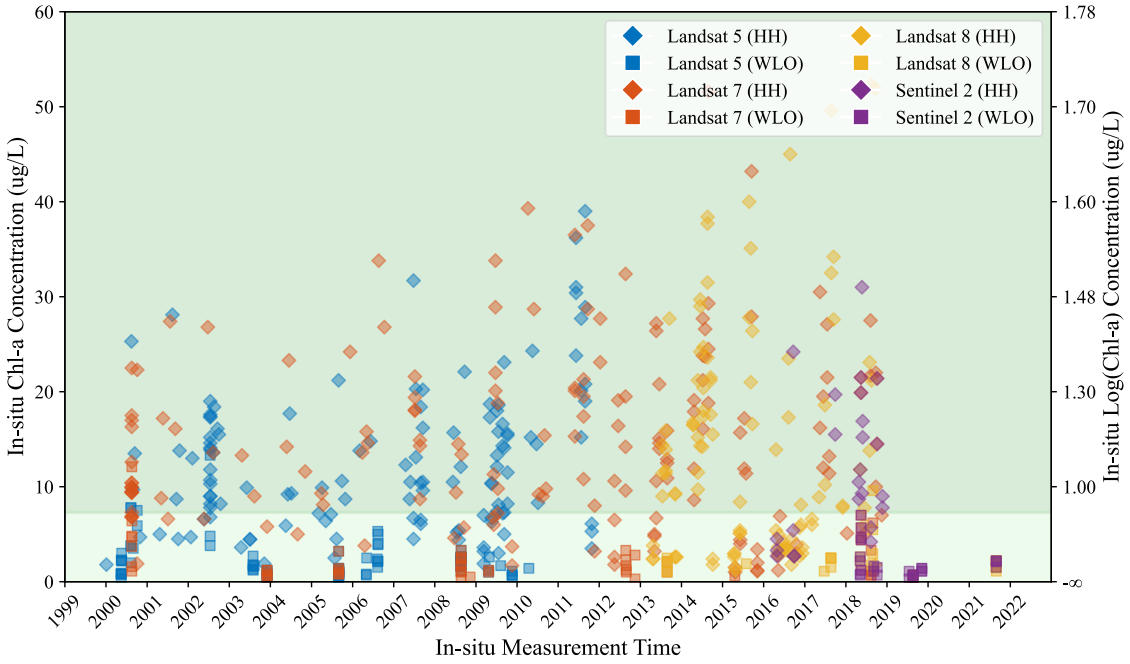

Figure A1.1. Time-series plot of in-situ data, displaying Hamilton Harbour (HH, diamond markers) and Western Lake Ontario (WLO, square markers) alongside Landsat 5 (blue), 7 (red), 8 (yellow), and Sentinel-2 (purple) matchups. The green background represents classes of oligotrophic/mesotrophic and eutrophic/hypereutrophic based on Carlson's Trophic State Index (TSI). Two outlier concentrations of 137 and 80 µg/L are excluded for better visual presentation.

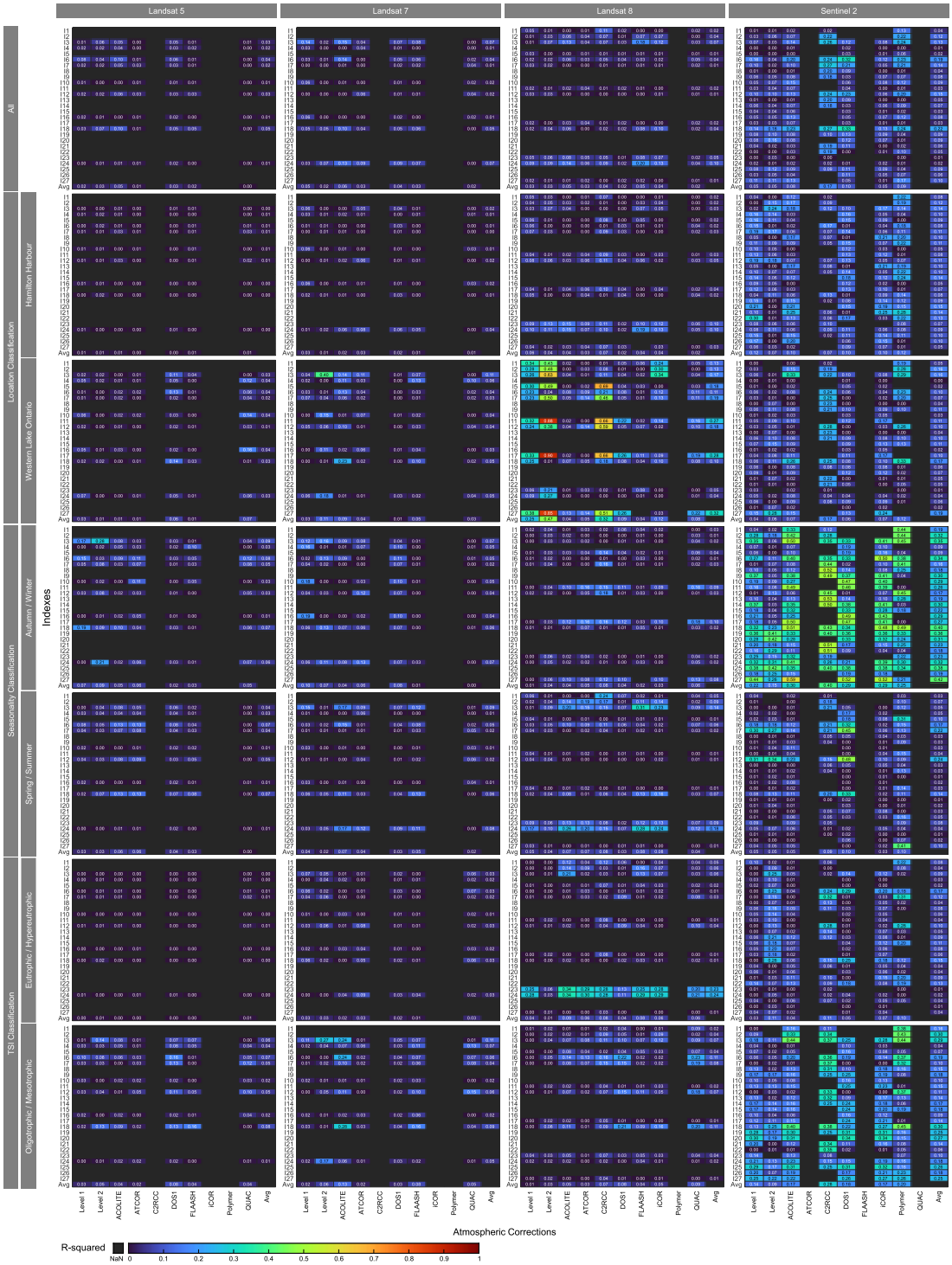

Figure A1.2. Heatmaps of R^2 between in-situ Chl-*a* concentration and corresponding index values of co-located pixels across various schemes and subcategories. Warmer colours indicate higher R^2 (better performance), while colder ones signify lower R^2. Black cells indicate N/A values.

Table A1.1. Definition of trophic classes from various studies.

		Trophic Classification				Reference
		Oligotrophic	Mesotrophic	Eutrophic	Hypereutrophic	
Chl-*a* (µg/L)	Range	0 – 3.3	3.3 – 8.3	8.3 – 27.5	27.5 <	(Shi, Zhang, Song, et al., 2019)*
		0 – 0.9	0.9 – 7.2	7.2 – 55.5	55.5 <	(M. Hu et al., 2021; Tuygun et al., 2023)*, †
		0 – 1.0	1.0 – 4.0	4.0 – 65.0	65.5 <	(Duan et al., 2008)*
		0 – 3.0	3.0 – 6.0	6.0 – 20.0	20.0 <	(Dove & Chapra, 2015)

* Based on Carlson's TSI (Carlson, 1977). † Used in this study.

Table A1.2. Descriptive statistics of the in-situ data categorized based on seasonality, study location and Carlson's Trophic State Index (TSI).

	Location Category			Seasonality Category		TSI Category	
	All	HH	WLO	Autumn/Winter	Spring/Summer	Oligotrophic/Mesotrophic	Eutrophic/Hypereutrophic
Count or 'n'	600	410	190	132	468	305	295
Frequency (%)	100%	68%	32%	22%	78%	51%	49%
Mean (µg/L)	10.5	14.3	2.2	9.6	10.7	2.9	18.3
Std. Deviation (µg/L)	11.3	11.8	2.0	10.0	11.7	1.9	11.6
Minimum (µg/L)	0.2	1.0	0.2	0.2	0.3	0.2	7.3
Maximum (µg/L)	137.0	137.0	13.3	51.9	137.0	7.2	137.0

Table A1.3. Atmospheric correction possessors utilized in this study. Level-2 products of Landsat 5 and 7 (LEDAPS- and LaSRC-corrected) are readily available for download, therefore excluded from the table below.

Name	Supported Sensors*	Input Products	Output Products	Processing Environment	Pros	Cons	Main Application	Processing Speed	Latest Version (Release Year)
ACOLITE	TM, ETM+, OLI, MSI	DN, TOA reflectance and ancillary data	TOA reflectance, surface level reflectance, normalized water-leaving reflectance, Rayleigh corrected reflectance, remote sensing reflectance, TSM concentration, turbidity, Chl-*a* concentration, and more	Python Graphical User Interface (GUI) and Python Command Line Interface (CLI)	Freely available, supports many sensors, flexible and customizable, user-friendly, glint correction, masking, Q/A forum support, batch mode	Limited to the processing of coastal and inland water	Aquatic	Moderate/Slow (depending on the selected outputs)	20221114.0 (2022)
ATCOR 3	TM, ETM+, OLI, MSI	DN, TOA reflectance, DEMs and ancillary data	TOA reflectance, ground reflectance, surface temperature, water vapour map	Geomatica (CATALYST), IDL, and Python	Supports many sensors, flexible and customizable, haze removal, topographic correction, cirrus removal, batch mode	Commercial, manual parameter selection and adjustment	Land, coastal, and water	Fast	9.4.0 (2022)
C2RCC	OLI, MSI	TOA reflectance and ancillary data	Surface reflectance, normalized water-leaving reflectance (ρ_{wn}), Chl-*a* concentration, TSM concentration and their related uncertainties and more	Plugin to the ESA SNAP	Freely available, designed for water, open-source code, based on machine learning and neural network, correction for many factors (gaseous absorption, air pressure, etc.)	Limited to a few sensors, unsuitable for non-water applications	Coastal and inland water	Moderate (depending on the selected outputs)	-
DOS1	TM, ETM+, OLI, MSI	DN, TOA Reflectance	Surface reflectance	Plugin to the QGIS (Semi-Automatic Classification Plugin, SCP)	Freely available, easy to implement, suitable for a wide range of sensors, open source	Simplifying assumptions, less accurate for complex atmospheric conditions	Land, coastal, and water	Fast	7.10.11 (2023)
FLAASH	TM, ETM+, OLI	Radiance in BIL or BIP format and ancillary data	Apparent reflectance (upscaled to 10,000), water vapour, cloud map (only for hyperspectral input)	Plugin to the ENVI, stand-alone C++ module	Reliable, supports a wide range of sensors, available for multispectral and hyperspectral sensors, water and aerosol retrieval and cloud masking	Commercial, manual parameter selection and adjustment	Land, coastal, and water	Slow	4.7 (2009)

iCOR (previously known as OPERA)	OLI, MSI	TOA reflectance	BOA reflectance	Plugin to the ESA SNAP	Freely available (basic version), generic, requires few input parameters, available for spaceborne, airborne and drone images, correction for adjacency effects	Limited sensor support, limited output products	Land, coastal, and water	Moderate	3.0.0
Polymer	MSI	TOA reflectance and ancillary data	Water reflectance and more	Python CLI	Freely Available Flexible and robust, Q/A forum support, correction for sunglint, batch mode, land mask, cloud mask	Unsuitable for non-water applications, user-unfriendly	Water	Moderate	4.16 (2023)
QUAC	TM, ETM+, OLI	DN, apparent reflectance or radiance	Apparent reflectance (upscaled to 10,000)	Plugin to the ENVI, stand-alone C++ module	Easy to use, supports a wide range of sensors, does not require ancillary data, does not require knowledge of sensor metadata, available for hyperspectral and multispectral imagery	Commercial, less accurate, unsuitable for water-only scenes	Land, coastal, and water	Fast	-
Sen2Cor	MSI	TOA reflectance and ancillary data	BOA reflectance	Plugin to the ESA SNAP and Python CLI	Freely available, reliable for MSI, correction for terrain and cirrus, and aerosol optical thickness, water vapour, scene classification map and quality indicators for cloud and snow probabilities	Limited to MSI	Land, coastal, and water	Moderate	2.11.00 (2022)

* Only limited to the list of sensors used in this study.

Abbreviations in alphabetical order: BOA: Bottom-of-Atmosphere, DN: Digital Number, TOA: Top-of-Atmosphere

Table A1.4. Summary of RS-derived Chl-*a* models across satellites and data categories. L5 = Landsat 5, L7 = Landsat 7, L8 = Landsat 8 and S2 = Sentinel-2

Model	**AC**	**RS-Derived Chl-*a***	**n**	**RMSE (µg/L)**	**RMSLE**	**Bias**	**MAE**	**MAPE (%)**	**MDAPE (%)**	**ε (%)**	**β (%)**	**R^2**
M_{L5-All}	ACOLITE	$Chla_{RS}$= -14.00×I_{18} + 11.85	201	11.25	0.41	1.53	2.12	139.00	59.80	87.45	39.15	0.35
M_{L5-HH}	QUAC	$Chla_{RS}$= 4.90×I_7 + 5.47	123	13.23	0.31	1.25	1.74	84.13	41.30	54.46	24.94	0.01
M_{L5_WLO}	QUAC	$Chla_{RS}$= 1.03×I_4 + 1.59	82	1.82	0.21	1.02	1.43	36.10	26.69	28.86	12.31	0.19
M_{L5-AW}	Level-2 (LEDAPS)	$Chla_{RS}$= -25.20×I_3 + 31.74	42	5.35	0.39	1.35	2.08	123.11	51.80	94.68	25.01	0.30

M_{L5-SS}	ATCOR	$Chla_{RS}$= -19.68×I_{18} + 14.74	161	12.15	0.39	1.44	1.99	126.12	50.65	73.14	33.14	0.38
M_{L5-EH}	QUAC	$Chla_{RS}$= 11.55×I_{12} + 13.89	93	14.12	0.21	1.15	1.46	43.68	37.41	41.45	12.30	0.00
M_{L5-OM}	DOS1	$Chla_{RS}$= -2.48×I_{6} + 7.53	112	1.68	0.22	1.08	1.54	48.03	42.24	56.64	12.38	0.25
M_{L7-All}	Level-2 (LEDAPS)	$Chla_{RS}$= -15.52×I_{3} + 26.41	216	9.64	0.49	1.56	2.48	202.12	61.20	107.78	24.78	0.10
M_{L7-HH}	ACOLITE	$Chla_{RS}$= 500.40×I_{24} + 10.26	160	8.99	0.35	1.26	1.80	109.47	41.75	52.45	10.17	0.04
M_{L7-WLO}	Level-2 (LEDAPS)	$Chla_{RS}$= -19.41×I_{3} + 21.53	57	1.51	0.20	1.02	1.47	40.61	34.53	38.17	11.35	0.38
M_{L7-AW}	Level-1	$Chla_{RS}$= -47.12×I_{4} + 28.64	39	10.12	0.52	1.72	2.68	210.25	69.72	189.76	59.91	0.19
M_{L7-SS}	Level-2 (LEDAPS)	$Chla_{RS}$= -10.74×I_{3} + 22.04	178	9.33	0.48	1.51	2.35	192.44	56.90	83.35	22.78	0.07
M_{L7-EH}	ATCOR	$Chla_{RS}$= 0.01×I_{24} + 17.57	122	7.75	0.18	1.08	1.42	38.36	31.73	40.00	9.08	0.08
M_{L7-OM}	Level-2 (LEDAPS)	$Chla_{RS}$= -15.79×I_{3} + 18.52	94	1.69	0.24	1.11	1.62	57.22	45.48	63.44	15.65	0.30
M_{L8-All}	FLAASH	$Chla_{RS}$= 0.08×I_{24} + 3.64	115	12.29	0.49	1.57	2.49	204.36	66.45	122.72	46.81	0.08
M_{L8-HH}	FLAASH	$Chla_{RS}$= 0.08×I_{24} + 5.66	96	12.67	0.44	1.44	2.21	167.45	60.91	100.66	26.44	0.08
M_{L8-WLO}	Level-1	$Chla_{RS}$= 23.72×I_{1} - 24.32	24	1.39	0.18	1.13	1.39	38.71	25.19	31.38	14.02	0.39
M_{L8-AW}	C2RCC	$Chla_{RS}$= 4.72×I_{7} - 0.67	32	11.33	0.46	1.53	2.38	175.36	65.65	102.25	53.15	0.14
M_{L8-SS}	FLAASH	$Chla_{RS}$= -27.09×I_{3} + 36.80	93	11.72	0.45	1.46	2.15	161.20	64.74	103.30	36.31	0.23
M_{L8-EH}	ACOLITE	$Chla_{RS}$= -41.56×I_{3} + 53.14	62	11.58	0.21	1.15	1.49	49.88	39.97	45.99	13.32	0.26
M_{L8-OM}	QUAC	$Chla_{RS}$= -0.80×I_{6} + 4.22	63	1.25	0.18	1.08	1.41	39.04	32.75	39.32	10.32	0.24
M_{S2-All}	DOS1	$Chla_{RS}$= -17.38×I_{18} + 9.50	51	6.19	0.38	1.36	2.00	119.66	54.24	76.54	39.85	0.36
M_{S2-HH}	Polymer	$Chla_{RS}$= 27.64×I_{2} - 0.14	24	6.97	0.31	1.24	1.79	85.43	44.74	68.26	15.53	0.16
M_{S2-WLO}	ACOLITE	$Chla_{RS}$= -5.84×I_{3} + 7.99	27	1.10	0.17	1.03	1.36	35.05	24.10	28.30	0.85	0.37
M_{S2-AW}	ACOLITE	$Chla_{RS}$= -17.57×I_{18} + 15.51	18	5.95	0.37	1.30	1.68	117.28	26.20	35.50	10.84	0.49
M_{S2-SS}	DOS1	$Chla_{RS}$= -57.12×I_{12} + 12.11	33	5.11	0.37	1.16	2.04	105.54	58.21	104.07	2.28	0.32
M_{S2-EH}	DOS1	$Chla_{RS}$= 22.01×I_{18} + 17.93	16	5.50	0.15	1.06	1.32	29.56	22.51	24.37	13.08	0.32
M_{S2-OM}	ACOLITE	$Chla_{RS}$= -5.01×I_{3} + 7.08	35	1.16	0.16	1.03	1.36	33.36	28.22	36.06	5.66	0.52

Appendix II

Supplementary Materials from Chapter 4

1. Quality Control of Remote Sensing Data using the Quality Assessment (QA) Band

Each Landsat 8 and 9 collection 2 L1TP scene bundle includes a QA band file, which aids in the initial characterization of features such as clouds, cloud shadow, ice/snow, water, and land in a scene. Generated through the CFMask algorithm, this band, formatted as an unsigned 16-bit COG image, holds quality statistics from the cloud mask and scene details, maintaining dimensions similar to the original L1TP scenes. The QA band needs conversion to a binary 16-bit format for a detailed understanding, with each bit depicting unique information and confidence levels about identified features. Table A2.1 provides descriptions of high-confidence flags used in Figure A2.1, which showcases distributions of these features across analyzed downloaded scenes.

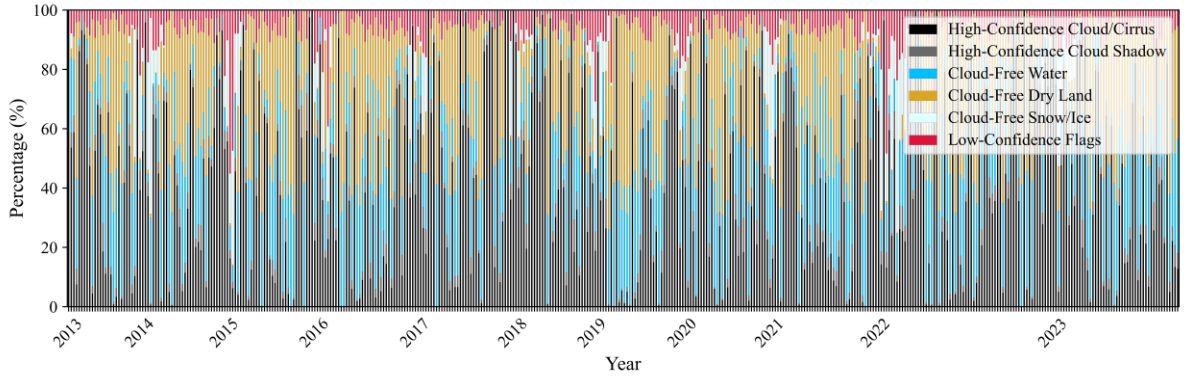

Figure A2.1. Percentage coverage of all high-confidence flags across all images, grouped by year from the QA band analysis

Table A2.1. Interpretation of the QA band pixel values for identifying high-confidence flags. (Department of the Interior USGS, 2022)

	High-Confidence Cloud/Cirrus			**High-Confidence Cloud Shadow**	**Cloud-Free Water**	**Cloud-Free Dry Land**	**Cloud-Free Snow/Ice**
QA Band Pixel Values	22280	55052	54724	23888	21952	21824	30048
Binary Value	01 01 01 11 0 0 0 0 1 0 0 0	11 01 01 11 0 0 0 0 1 1 0 0	11 01 01 01 1 1 0 0 0 1 0 0	01 01 11 01 0 1 0 1 0 0 0 0	01 01 01 01 1 1 0 0 0 0 0 0	01 01 01 01 0 1 0 0 0 0 0 0	01 11 01 01 0 1 1 0 0 0 0 0
Fill (Bit: 0)	Image data	Image data	Image data	Image data	Image data	Image data	Image data
Dilated Cloud (Bit: 1)	Cloud is not dilated or no cloud	Cloud is not dilated or no cloud	Cloud is not dilated or no cloud	Cloud is not dilated or no cloud	Cloud is not dilated or no cloud	Cloud is not dilated or no cloud	Cloud is not dilated or no cloud
Cirrus (Bit: 2)	Cirrus confidence: no confidence level set or low confidence	High confidence cirrus	High confidence cirrus	Cirrus confidence: no confidence level set or low confidence	Cirrus confidence: no confidence level set or low confidence	Cirrus confidence: no confidence level set or low confidence	Cirrus confidence: no confidence level set or low confidence
Cirrus Confidence (Bit: 14-15)	Low confidence	High confidence	High confidence	Low confidence	Low confidence	Low confidence	Low confidence
Cloud (Bit: 3)	High confidence cloud	High confidence cloud	Cloud confidence is not high	Cloud confidence is not high	Cloud confidence is not high	Cloud confidence is not high	Cloud confidence is not high
Cloud Confidence (Bit: 8-9)	High confidence	High confidence	Low confidence	Low confidence	Low confidence	Low confidence	Low confidence
Cloud Shadow (Bit: 4)	Cloud shadow confidence is not high	Cloud shadow confidence is not high	Cloud shadow confidence is not high	High confidence cloud shadow	Cloud shadow confidence is not high	Cloud shadow confidence is not high	Cloud shadow confidence is not high
Cloud Shadow Confidence (Bit: 10-11)	Low confidence	Low confidence	Low confidence	High confidence	Low confidence	Low confidence	Low confidence

Snow (Bit: 5)	Snow/Ice confidence is not high	Snow/Ice confidence is not high	Snow/Ice confidence is not high	Snow/Ice confidence is not high	Snow/Ice confidence is not high	Snow/Ice confidence is not high	High confidence snow cover
Snow/Ice Confidence (Bit: 12-13)	Low confidence	Low confidence	Low confidence	Low confidence	Low confidence	Low confidence	High confidence
Clear (Bit: 6)	Cloud or dilated cloud bits are set	Cloud or dilated cloud bits are set	Cloud and dilated cloud bits are not set	Cloud and dilated cloud bits are not set	Cloud and dilated cloud bits are not set	Cloud and dilated cloud bits are not set	Cloud and dilated cloud bits are not set
Water (Bit: 7)	Land or cloud	Land or cloud	Water	Land or cloud	Water	Land or cloud	Land or cloud

A similar analysis is conducted on the atmospherically corrected ($\boldsymbol{\rho_w}$) images. Figure A2.2 illustrates the distribution of available water pixels within the region of interest for both groups of Landsat scenes: the 17/30 and 18/30 path/row configurations.

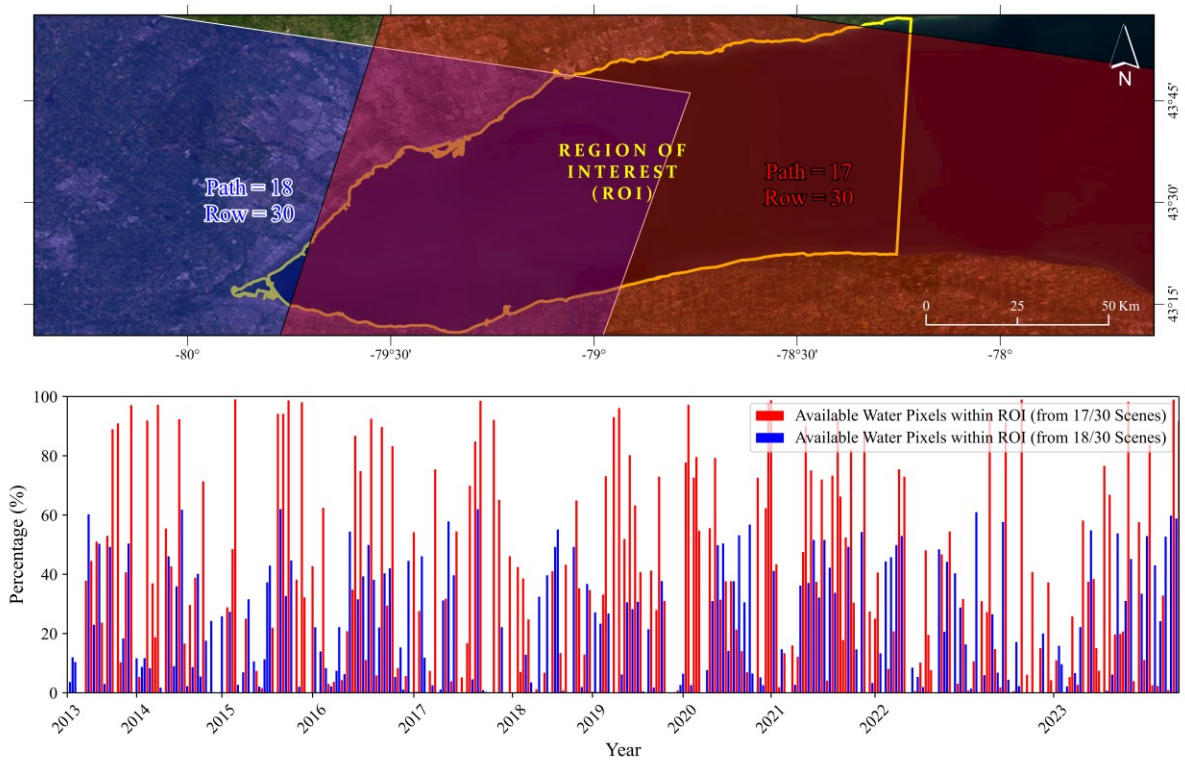

Figure A2.2. Water pixel availability of collected imagery

2. Atmospheric Correction using ACOLITE (Further Details)

Minor adjustments have been made to the default settings of ACOLITE to ensure high-quality output products of ρ_w, effectively correcting for atmospheric effects. These adjustments are as follows: The non-water masking threshold, $\rho(1610)$ band, has been raised from 0.0215 to 0.050; the cirrus masking threshold, from 0.005 to 0.010; and the TOA reflectance masking threshold, from 0.3 to 0.5 for a more conservative masking process of non-water pixels. Additionally, the Gaussian smoothing kernel size for the L2W mask has been increased from the default 3 to 4. Furthermore, cirrus correction using Landsat's $\rho(1370)$ band, as well as control of residual glint correction, have been implemented. Also, EARTHDATA credentials have been provided for accessing ancillary and SRTM DEM data. This has resulted in high-quality water-leaving products apt for a Chl-*a* retrieval study. Upon inspecting the pixel

values of Level-1 images (before atmospheric correction) and ACOLITE-processed images, a significant difference and improved correlation of the spectral signatures with recorded field Chl-*a* measurements of the same pixels can be observed (see Figure A2.3).

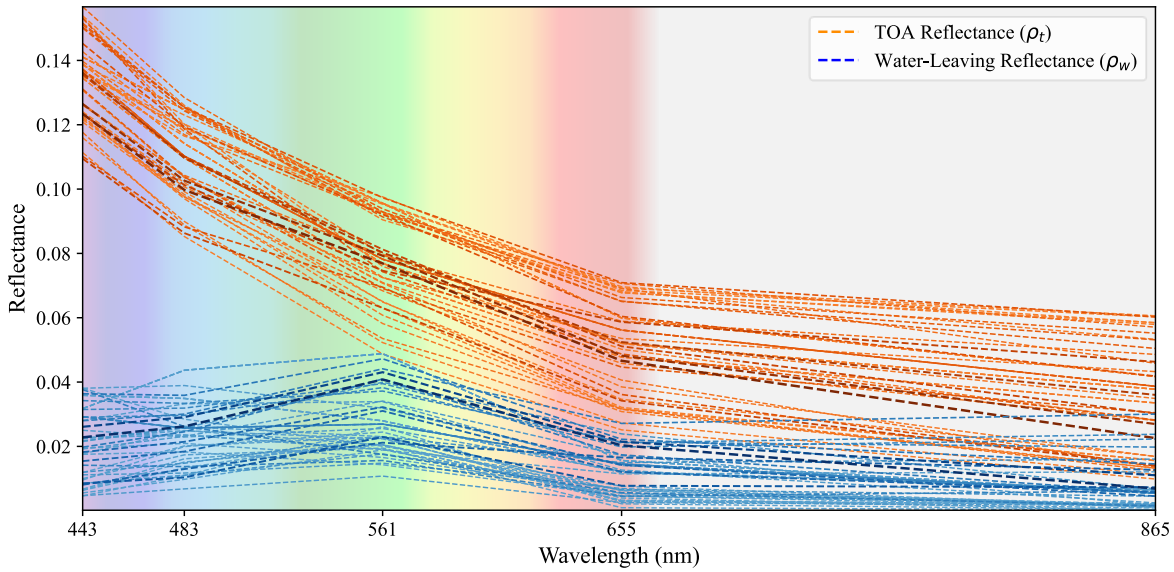

Figure A2.3. Comparing the matchup pixel reflectance's before ($\boldsymbol{\rho_{TOA}}$) and after ($\boldsymbol{\rho_w}$) atmospheric correction. The thickness of the dashed lines positively corresponds to their in-situ Chl-*a* concentration. The background colour scheme depicts the visible spectrum.

3. Chl-*a* Retrieval Model (Further Details)

In order to better approximate a normal distribution, the in-situ Chl-*a* data underwent a logarithmic (base 10) transformation. Figure A2.4 illustrates the results of the feature importance analysis, with (a) displaying the simple Spearman's and Pearson's correlation coefficients and (b) reporting the SHAP value, both highlighting the highest importance of I_3 in comparison to other reflectance indexes.

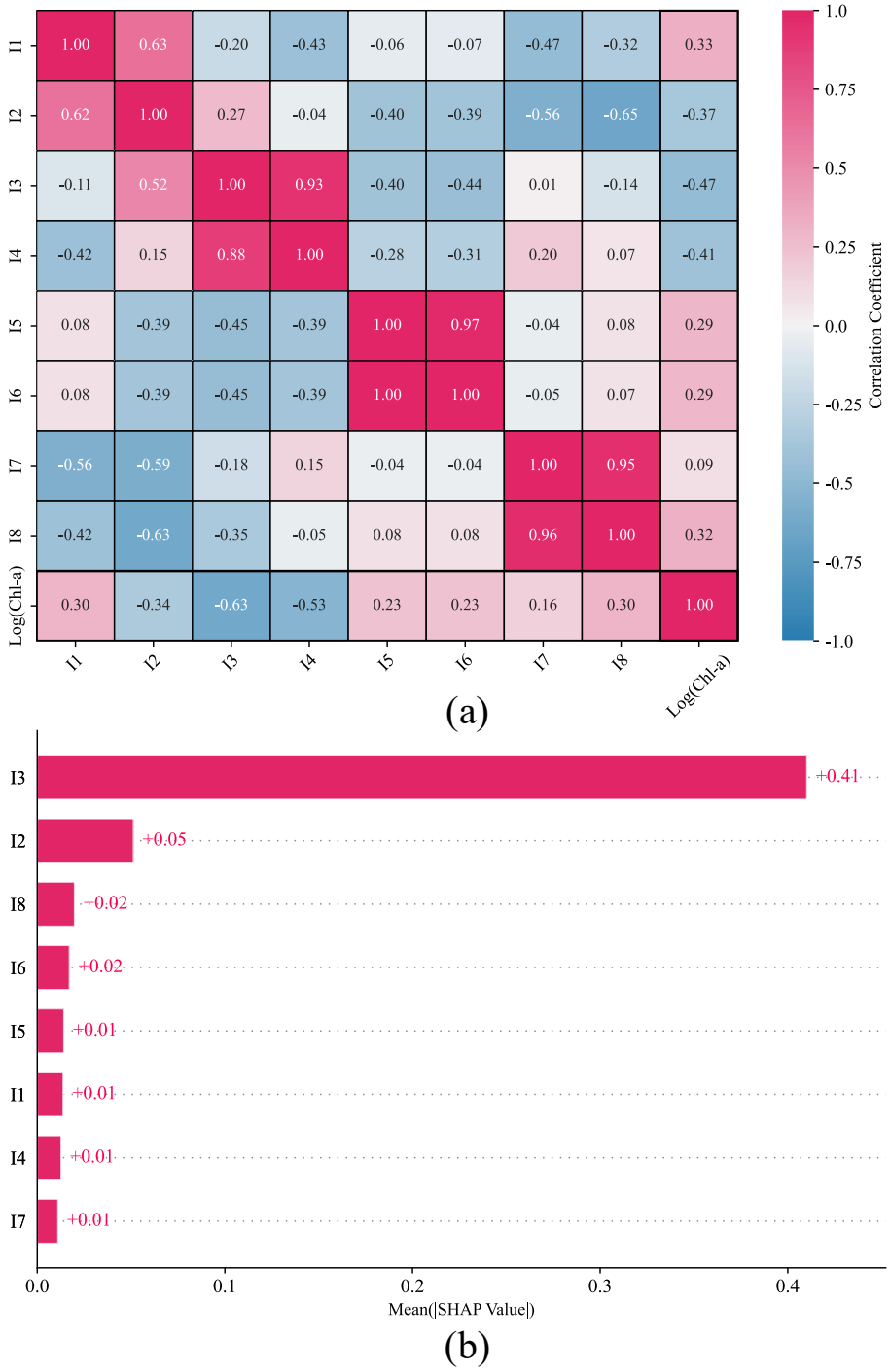

Figure A2.4. Results of the feature importance analysis. Figure (a) illustrates the Pearson (top triangle) and Spearman's rank (bottom triangle) correlations coefficients. Figure (b) depicts the SHAP value for each feature indicating their importance in predicting Chl-*a*.

After selecting the input feature (I_3) for the regression model, the fitted trendline, based on an exponential trend, is selected with RMSLE of 0.24 and R^2 of 0.41, as illustrated in Figure A2.5 (b). The resulting regression equation is then used to generate 'Modeled Chl-*a*' values, which are plotted against the 'In-Situ Chl-*a*,' showcasing the overall performance of the model in Figure A2.5 (a). The performance metrics of RMSE, MAE, MAPE of 8.49, 5.06, and 78% were calculated (averaged) respectively, using a 5-fold cross-validation, allocating 80% of the data for training and the remaining 20% for testing in each iteration.

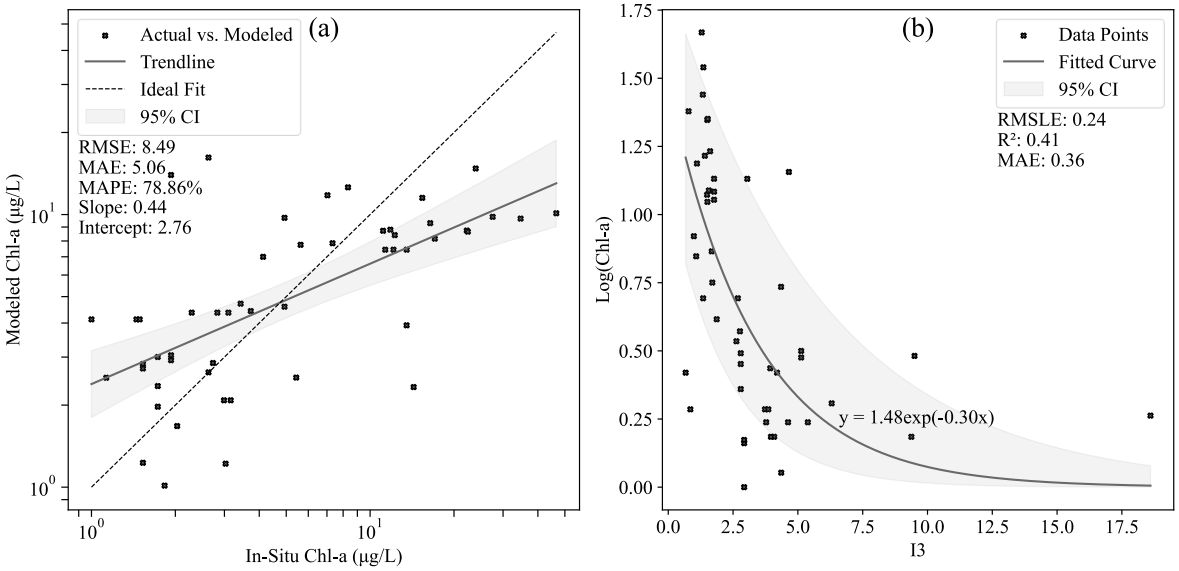

Figure A2.5. Results of the regression analysis. Figure (a) illustrates the measured vs model Chl-*a* concentration relationship. Figure (b) depicts the relationship between the predictor (I_3) and response (Chl-*a*)

4. Spatial-Temporal Variability and Trends (Further Details)

Figure A2.6 presents annual and monthly boxplots for all available in-situ data from 2013 to 2023 for Hamilton Harbour, with maximum and average points colour-coded to enhance visibility. Despite the presence of numerous extreme events – highlighted particularly in 2014, 2016, 2017, and the months of April and October, with the exceptions being September and February 2021 which did not exhibit any outliers – alongside significant data gaps in 2019 and 2020, and a scant number of records in January, February, and December of 2022, this dataset was instrumental in validating the bloom intensity results for Hamilton Harbour.

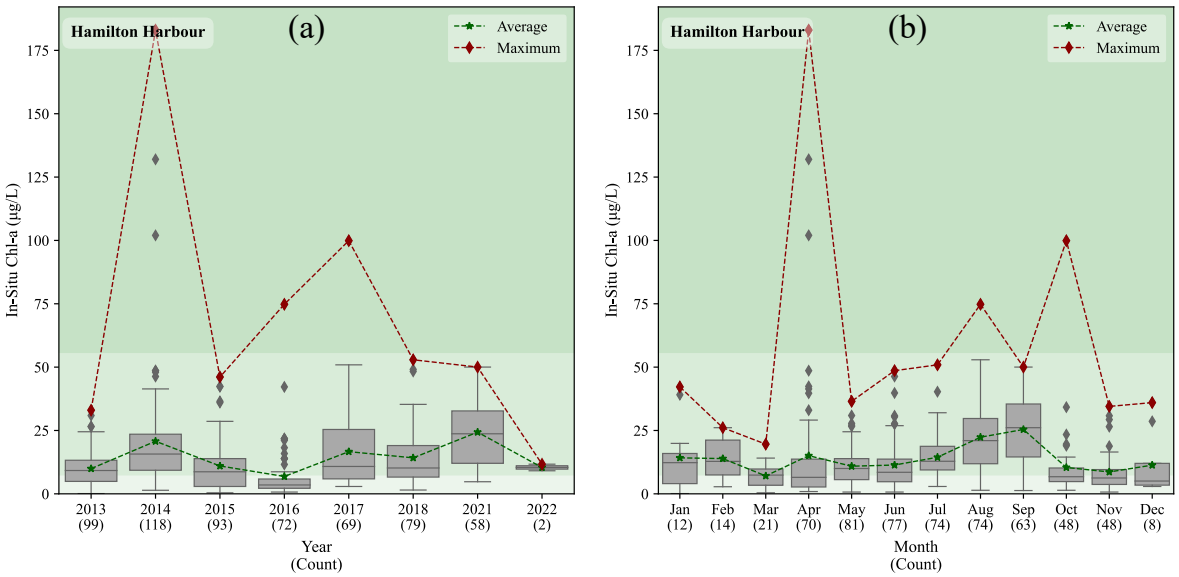

Figure A2.6. Annual and monthly boxplots of in-situ Chl-*a* concentration data from Hamilton Harbour, compiled from various sources for the period 2013-2023. The green and red lines indicate the average and maximum values, respectively, for each (a) year and (b) month. The shaded green background denotes different trophic states based on Carlson's TSI: oligo-mesotrophic (up to 7.2 µg/L), eutrophic (7.2 to 55.5 µg/L), and hypereutrophic (more than 55.5 µg/L)

For Lake Ontario, due to the insufficiency and spatial distribution of sampling points, we opted for the NOAA dataset comprising satellite-derived surface Chl-*a* concentration estimates from the MODIS sensors, with spatial resolution of 250 m. NOAA's dataset utilizes a band-ratio retrieval algorithm, based on GLNPO monitoring data, to offer monthly averages that highlight long-term trends and seasonal variations in Chl-*a* concentration across the Great Lakes. Our analysis calculated the average pixel values from annual average maps, incorporating error bars for variability, and plotted these against NOAA's average estimated Chl-*a* concentration (see Figure A2.7) accessible through the NOAA's National Marine Ecosystem Status website. A key distinction is that our data focus on the western part of Lake Ontario, whereas NOAA's dataset represents lake-wide averages derived from MODIS.

The results of the multidimensional trend analysis conducted by ArcGIS Pro are displayed in Figure A2.8 for (a) the slope and (b) RMSE of a linear trend, and (c) Sen's slope of the Mann-Kendall test. Sen's slope, a non-parametric measure, provides a robust estimate of the median trend slope, indicating the magnitude and direction of trends over time. Here, (a) and (c) indicate the strength and direction of the trend, while (b) highlights the uncertainties associated with it.

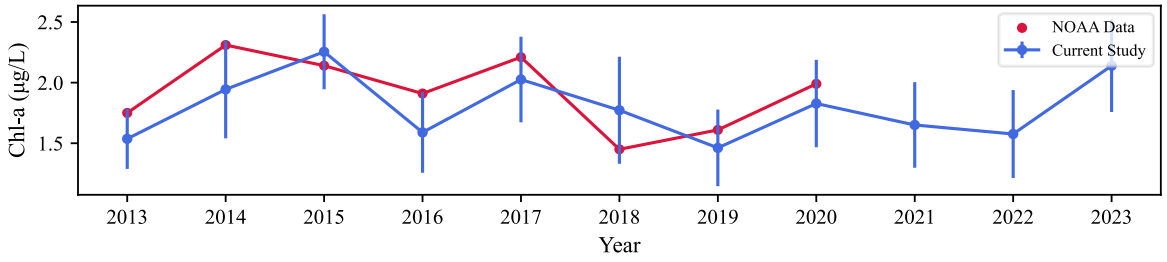

Figure A2.7. Comparison of satellite-derived annually averaged lake-wide estimates of Chl-*a* concentration reported in NOAA dataset (red) versus our maps (blue), from 2013 to 2023. The error bars represent standard deviations from the measurements.

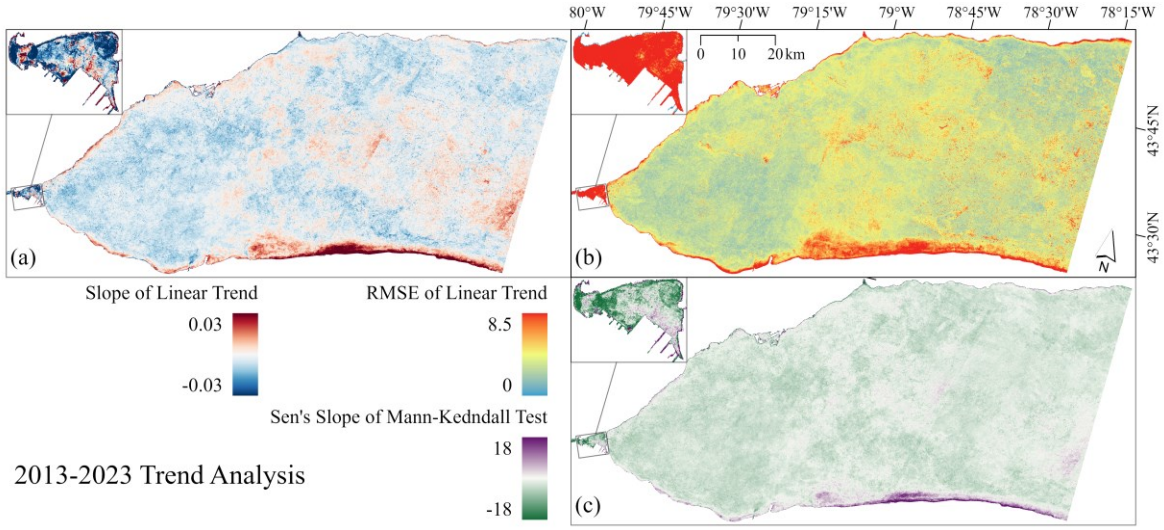

Figure A2.8. Maps of trend analysis of 2013-2023 estimated Chl-*a* concentration maps: (a) slope of linear trend (b) RMSE of linear trend (c) Sen's slope of Mann-Kendall test

The pixel values of the annual (a) and monthly (b) averaged estimated Chl-*a* concentration maps are plotted against their shortest distance from the shore in Figure A2.9, colour-coded based on different years or months, showcasing an exponential-like distribution with a distinct threshold at 3 km.

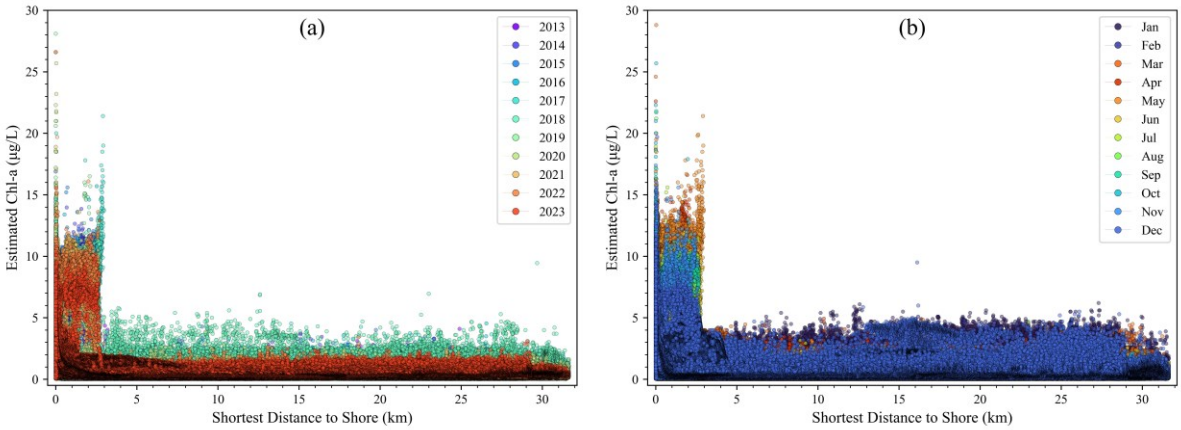

Figure A2.9. Distribution of estimated Chl-*a* concentration in relation to distance from the shoreline, for (a) the annual and (b) the eleven-year monthly averages

Glossary

Algal Blooms	Algal blooms are defined by a significant increase in algal biomass. Depending on the context and field, the definition might differ. Various factors could be taken into account, for example, size and community composition of algae. Here, algal blooms are defined as Chl-*a* concentration above 10 µg/L, which aligns with the World Health Organization's guidelines of exceeding 20,000 cyanobacterial cells per mL.
Apparent Optical Properties (AOPs)	AOPs are optical measurements that depend on both the IOPs of the water and the environmental conditions, such as the angle and quality of incident sunlight and the viewing geometry of the sensor. AOPs, such as water-leaving reflectance and remote sensing reflectance, are essential for remote sensing applications as they are directly observable quantities that help in assessing water quality and other aquatic characteristics.
Apparent Reflectance	Apparent reflectance is derived when radiance is converted to reflectance, a value that embodies both the spectral properties of the material under observation and the methodology utilized in the conversion. It accounts for external influences such as atmospheric conditions and viewing angles, making it dependent on the specific circumstances of observation rather than being a fundamental property of the object.
Atmospheric Correction	Atmospheric correction is a procedure in remote sensing to eliminate the effects of the atmosphere on the imagery data. By mitigating the influences of absorption and scattering phenomena, it retrieves ground reflectance values that are representative of the true characteristics of the earth's surface.
Case I and Case II Waters	Case 1 and Case 2 Waters are classifications used in aquatic remote sensing to distinguish between different types of water bodies based on their optical properties. Case 1 Waters are primarily influenced by algae and related organic materials, where the absorption and scattering of light are mostly governed by chlorophyll concentration. Case 2 Waters, on the other hand, are characterized by a more complex mix of substances, including sediments, organic matter, and pollutants, which significantly influence their optical properties.

Cyanobacteria	Cyanobacteria, often referred to as blue-green algae (BGA), are microorganisms that occur naturally in aquatic environments. They flourish in warm, still or slow-moving waters that have a high nutrient content and get abundant sunlight. Microcystins are the most commonly found toxins produced by cyanobacteria. These invisible, tasteless, and odorless toxins present hazards to the well-being of humans and animals.
Digital Number (DN)	DN signifies the pixel values in raw images, which are derived from the quantifiable amounts of energy captured by a sensor. These values are recorded in pixels, creating digital images, with a range determined by the sensor's radiometric resolution. For instance, Landsat 8 has a 16-bit resolution, which produces pixel values between 0 to 65536, where higher values denote greater radiant energy.
Inherent Optical Properties (IOPs)	IOPs are the fundamental optical properties of water that are independent of the ambient light field. They are determined by the water itself and the materials dissolved or suspended within it, such as phytoplankton, detritus, and dissolved organic matter. The key IOPs include absorption and scattering coefficients, which influence how light penetrates, gets absorbed, and is scattered within the water body.
Irradiance	Irradiance refers to the flux of radiant energy received per unit area of a surface, usually expressed in W/m^2 or W/cm^2. It measures the power of electromagnetic radiation incident on a surface per unit area, serving to quantify the intensity of radiant flux impacting a specific area.
Levels (Processing Levels)	Most satellite data are generally categorized into four distinct processing levels. Level 0: Unprocessed, raw satellite data. Level 1: Data corrected for geometric and radiometric errors, showing reflectance at the top of the atmosphere. Level 2: Data adjusted for surface reflectance, including atmospheric corrections. Level 3: Final data products (such as Chl-*a* maps), derived from level 2 data.

Normalized Water-Leaving Reflectance ($\boldsymbol{\rho_{wn}}$or $[\boldsymbol{\rho_w}]_{\boldsymbol{n}}$)	Similar to ρ_w but when the view angle is corrected so the values are as they would have been observed in the nadir. Most of the time it is pretty close to that of ρ_w.
Optically Active Constituents (OACs)	OACs refer to the substances in water bodies that influence its optical properties. These include phytoplankton pigments, suspended particulate matter, and CDOM.
Optical Water Types (OWTs)	OWTs refer to the classification of water bodies based on their optical properties as observed through remote sensing techniques. This categorization is determined by factors such as water colour, clarity, and the presence of substances like sediments and algae.
Radiance ($\boldsymbol{L}$)	Radiance (L) designates the amount of radiant energy perceived per unit area from a specific direction per unit solid angle, typically expressed in units of $Wm^2sr^{-1}\mu m^{-1}$. It is influenced by several factors including the reflectance properties of the target, adjacent pixels, atmospheric interactions, and radiation sources like the sun. Landsat images are supplied in DN values, which can be transformed into radiance using the equation: $$L(\lambda) = [Gain \times DN(\lambda)] + Offset$$
Radiometric Calibration	Radiometric calibration is the process of transforming DN values in remote sensing imagery into physical units such as radiance or reflectance. It accounts for sensor peculiarities, acquisition conditions, and calibration parameters, ensuring uniformity and comparability between different images and across various sensors.
Rayleigh Corrected Reflectance ($\boldsymbol{R_{RC}}$)	Rayleigh corrected reflectance is used in remote sensing to adjust satellite imagery for atmospheric scattering, specifically Rayleigh scattering. This process enhances the accuracy of surface reflectance data by removing the blue light scattering caused by atmospheric particles.

Reflectance ($\boldsymbol{\rho}$ or $\boldsymbol{R}$)	Reflectance (ρ), is a dimensionless ratio that illustrates the proportion of incident light that is reflected by a surface target. It is utilized in remote sensing to discern materials based on their distinct spectral signatures, offering a means to identify and analyze different materials and surfaces. It is worth noting that antithetical to reflectance are 'absorbance' and 'transmittance', representing the fraction of incident light absorbed and transmitted by the surface, respectively.
Remote Sensing Reflectance ($\boldsymbol{R_{RS}}$)	Remote Sensing Reflectance (R_{RS}) is a key parameter in aquatic RS. It is defined as the ratio of the water-leaving radiance [$L_W(\lambda)$] to the downwelling irradiance just above the surface of the water [$E_d(\lambda)$]. Mathematically, it is expressed as: $$R_{RS}(\lambda) = \frac{L_W(\lambda)}{E_d(\lambda)}$$
Surface Reflectance ($\boldsymbol{\rho_S}$ or $\boldsymbol{R_S}$)	Surface Reflectance (ρ_S) delineates the fraction of incident light that is reflected by the Earth's surface, devoid of atmospheric influences. It is derived by adjusting top-of-atmosphere reflectance for atmospheric effects, sensor attributes, and viewing geometry, providing a more accurate depiction of the Earth's surface properties.

Top-of-Atmosphere Reflectance ($\boldsymbol{\rho_{TOA}}$ or $\boldsymbol{R_{TOA}}$)	To diminish variations between different scene captures, images rendered in radiance can be converted into TOA reflectance. This metric integrates both surface and atmospheric reflections and accounts for solar irradiance to ensure consistency across various observations. This technique primarily serves to reduce between-scene variability by normalizing for solar irradiance, facilitating a uniform approach to the analysis of remote sensing imagery. The unitless parameter equates to the ratio of the reflected energy to the incoming solar power, and is derived through the formula: $$\rho_{TOA}\,(\lambda) = \frac{\pi \times L_{TOA}(\lambda) \times d^2}{ESUN(\lambda) \times \cos\,(\theta)}$$ where: d is Earth-Sun distance in astronomical units (AU) $L_{TOA}(\lambda)$ is the radiance measured at the top of the atmosphere by a satellite sensor $ESUN(\lambda)$ is the mean solar exoatmospheric irradiance (in $Wm^2\mu m^{-1}$) θ is the solar zenith angle (in degrees) Sentinel 2 images are provided in scaled TOA reflectance.
Water-Leaving Reflectance ($\boldsymbol{\rho_w}$)	Water-leaving radiance reflectance (or commonly known as water-leaving reflectance although this is somewhat of a misnomer) is a key AOP in the RS of water quality, paramount for retrieving information on OACs such as Chl-*a*. It is a measure of the light that has interacted with the water column and is reflected back towards the atmosphere and detected by a sensor. $$\rho_w(\lambda) = \pi \frac{L_w(0^+, \lambda)}{E_d(0^+, \lambda)} = \pi R_{RS}(\lambda)$$ Where $L_w(0^+, \lambda)$ and $E_d(0^+, \lambda)$ are the water-leaving radiance and the downward irradiance just above the surface respectively, and R_{RS} (sr^{-1}) is the remote sensing reflectance.