UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LARGE-SCALE PATTERNS AND REGULATION OF ECOSYSTEM METABOLISM AND DISSOLVED ORGANIC CARBON IN CANADIAN LAKES

THESIS

PRESENTED

AS A PARTIAL REQUIRMENT

OF THE DOCTORAT OF BIOLOGY

BY

AMIR REZA SHAHABINIA

MAY 2025

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PATRONS À GRANDE ÉCHELLE ET RÉGULATION DU MÉTABOLISME ÉCOSYSTÈMIQUE ET DU CARBONE ORGANIQUE DISSOUS DANS LES LACS CANADIENS

THÈSE

PRÈSENTÈE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN BIOLOGIE

PAR

AMIR REZA SHAHABINIA

MAI 2025

ACKNOWLEDGEMENTS

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Paul del Giorgio. Paul, your unwavering support, patience, guidance, and thoughtful feedback have shaped not only the direction of this thesis but also my growth as a scientist and as a person. Your mentorship has been a constant source of inspiration. I have learned so much from you, about both science and life, and I will carry those lessons with me always. You have been more than a supervisor; you have been a steady presence, like family, like a father figure during these years. I am deeply thankful for your trust and kindness.

I would like to express my sincere gratitude to the funding agencies that made this research possible—NSERC and Hydro-Québec. Their support of the del Giorgio lab and the CarBBAS and LakePulse research programs enabled me to pursue my studies free from financial uncertainty. I am truly thankful for this invaluable opportunity.

I am especially grateful to Yannick Huot, Lead PI of the NSERC-funded LakePulse project, for envisioning and leading this remarkable pan-Canadian research network. It has been an honor to contribute, even in a small way, to such a visionary and impactful initiative. A warm thank you to Maxime Fradette, whose help with GIS-based analyses and technical advice was essential to this work. I deeply appreciate your patience, expertise, and willingness to assist whenever I needed support.

To my co-authors Matthew Bogard, Jean-François Lapierre, and Ryan Hutchins—thank you for your insightful feedback, mentorship, and generosity with your time and knowledge. Your guidance has left a lasting impact on both my research and my growth as a scientist.

I am also sincerely thankful to Alice Parkes, whose dedication in the lab and behind-the-scenes support helped keep this research moving forward. Your efforts in coordinating logistics and lab work made a big difference.

Many thanks to Katherine Velghe and Maryline Robidoux, research associates at GRIL, for their kindness, expertise, and hands-on support with various lab analyses. I am truly grateful for the time and care you put into the lab work that supported this research.

I would also like to thank Mathilde Couturier for her assistance with PARAFAC analysis, and Masumi Stadler for generously sharing her knowledge of R—your guidance helped me through more than one difficult coding challenge. Many thanks to Candice Aulard and Flora Mazoyer for her kind help in translating the abstract of this thesis into French—your support is truly appreciated.

To Mario Muscarella, Pascal Bodmer, Michaela de Melo, Maximilian Lau, Pedro Maia Barbosa, and Sara Soria Píriz—thank you for the thoughtful conversations, critical feedback, and kind encouragement you offered during various stages of my PhD. Each of you contributed to this work more than you know.

Finally, I want to extend my warmest thanks to the incredible community of colleagues and friends who made this journey so rewarding. Working alongside Marie Gérardin, Martin Demers, Gabriel Duran, Joan Pere Casas-Ruiz, Sofia Baliña, Bianca Rodríguez-Cardona, Mariana Peifer Bezerra, Camille St-Arneault Sergerie, Caroline Fink-Mercier, Sara Mercier-Blais, Teresa Aguirrezabala, Cynthia Soued, Karelle Desrosiers, Tonya DelSontro, Shoji Thottathil and Marie-Pier Hébert has been a true privilege. Thank you for the inspiration and support that made even the most challenging moments worthwhile. To each of you: thank you very much.

Thank you to Bruno Cremella, Yudhistir Reddy, and Facundo Smufer for your friendship, support, and all the fun we shared along the way. Your companionship made the hard days lighter and the good days even better. I'm truly grateful for everything—from the laughter to the late-night talks—and for being there to help me get through this journey.

I would like to thank my family for their unwavering love and support throughout this journey. To my mom, dad, sister, and brother—your constant encouragement and belief in me helped carry me through every challenge. Thank you for your understanding, your patience, and for accepting the distance that came with my being in another country, far from home. I truly could not have done this without you.

A very special thank you to my amazing wife, Fereshteh. Your love, encouragement, and strength have been the anchor of my life throughout these years. You've given so much of yourself to support me, always with grace, patience, and endless care. Your belief in me, even during the toughest moments, kept me going. Life has been so much brighter, fuller, and more joyful since you became a part of it. I truly couldn't have done this without you—this accomplishment is just as much yours as it is mine.

DEDICATION

I dedicate this work to my wife: Fereshteh.

I Love you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AIC: Akaike information criterion.

AImod: Modified aromaticity index.

ANCOVA: Analysis of covariance.

ANOVA: Analysis of variance.

BIX: Biological index.

C: Carbon.

CDOM: Colored dissolved organic carbon.

***Chla*:** Chlorophyll.

DIC: Dissolved inorganic carbon.

d-excess: Deuterium excess.

DOC: Dissolved organic carbon.

DOM: Dissolved organic matter.

EEM: Emission excitation matrix.

ESI: Electrospray ionization.

FDOM: Fluorescent dissolved organic matter.

FT-ICR MS: Fourier-transform ion cyclotron resonance mass spectrometry.

GPP: Gross primary production.

GPP/R: Gross primary production/Respiration.

HIX: Humification index.

HPLC–SEC: High-performance liquid chromatography–size exclusion chromatography.

L-CARE: Landscape carbon accumulation through reduction in emissions.

MAE: Mean absolute error.

MAPE: Mean absolute percentage error.

MF: Molecular formulae.

NEP: Net ecosystem production.

NLA: National lake assessment.

NMDS: Non-metric multidimensional scaling.

NMR: Nuclear magnetic resonance.

NOSC: Nominal oxidation state of carbon.

NSERC: Natural Sciences and Engineering Research Council of Canada.

PAR: Photosynthetically active radiation.

PARAFAC: Parallel-factor analyses.

PCA: Principal component analysis.

PC: Principal component.

PLS: Partial least square.

R: Respiration.

RDA: Redundancy analysis.

RMSE: Root-mean-square error.

SEM: Structural equation model.

SOC: Soil organic carbon

SR: Spectral slope ratio.

SUVA: Specific ultraviolet absorbance.

TN: Total nitrogen.

TP: Total phosphorus.

Turkey's HSD: Tukey's honestly significant difference test.

WRT: Water residence time.

RÉSUMÉ

Le métabolisme lacustre ainsi que la concentration et la composition de la matière organique dissoute (MOD) jouent des rôles cruciaux dans le façonnage des processus écosystémiques et la dynamique du carbone (C) dans les lacs. Malgré leur importance dans les cycles biogéochimiques, ma compréhension de leurs patrons à grande échelle et de leur régulation était jusqu'alors limitée à des bassins versants spécifiques et à un petit nombre de lacs, souvent le long de gradients environnementaux réduits. De plus, les liens directs potentiels entre le métabolisme du lac et la composition de la MOD sont encore mal compris. J'ai mené une étude à échelle continentale du métabolisme lacustre et de la MOD — tant sur le plan de la concentration que sur celui de la composition — dans des lacs du Canada, dans le cadre du Réseau LakePulse. Dans le chapitre 1, j'ai exploré les patrons spatiaux et les facteurs environnementaux relatifs à la production primaire brute (PPB), à la respiration (R) et à la production nette de l'écosystème (PNE) dans les lacs canadiens. Ensuite, j'ai mesuré diverses propriétés moléculaires et optiques de la MOD et examiné comment ces propriétés se rapportent aux variables lacustres, climatiques et hydrologiques (chapitre 2). Finalement, j'ai tenté de démêler et de comprendre les relations complexes qui existent entre la composition de la MOD et le métabolisme du lac (chapitre 3).

Dans le chapitre 1, j'ai décrit les patrons du métabolisme lacustre à travers le Canada, en mesurant la PPB, la R et la PNE pendant l'été. En utilisant un isotope de l'oxygène ($\delta^{18}O_2$), j'ai évalué les taux métaboliques dans 742 lacs répartis dans les principaux bassins versants et dans une diversité de paysages. La PPB et la R variaient considérablement, principalement en fonction des concentrations en nutriments, du carbone organique dissous (COD) et de la chlorophylle. La PNE présentait des liens faibles mais significatifs avec la disponibilité de la lumière et la matière organique dissoute colorée (MODC). Mes résultats révèlent des différences régionales dans le métabolisme de base et suggèrent que la sensibilité des lacs à l'eutrophisation et au brunissement est influencée par l'hydrologie et le climat à l'échelle du paysage.

Dans le chapitre 2, j'ai examiné comment différentes variables relatives au lac, au bassin versant et au climat influencent la concentration et la composition de la MOD dans les lacs du Canada. J'ai analysé les caractéristiques du COD et de la MOD dans 548 lacs répartis dans cinq grands bassins versants, en utilisant des mesures optiques, une analyse factorielle parallèle et de la spectrométrie de masse à très haute résolution. La concentration et la composition de la MOD variaient à la fois au sein des régions et entre elles, et présentaient des liens étroits avec l'hydrologie (estimée par l'excès de deutérium (d-excess)) et

l'état trophique du lac (estimé par le phosphore total (TP)). J'ai trouvé des relations régionales spécifiques entre les classes moléculaires individuelles et l'excès de deutérium, ce qui met en évidence des contrastes à l'échelle régionale dans la dynamique de la MOD. Bien qu'il existe des spécificités régionales dans la composition de la MOD reconstruite à partir des sources, les processus de transformation dans le réseau hydrologique sont apparus comme étant le principal facteur expliquant les variations d'une région à l'autre. Dans l'ensemble, mes résultats soulignent que ces processus jouent un rôle plus important que l'origine des apports de la MOD dans le façonnage de la composition de la MOD lacustre à l'échelle continentale.

Dans le chapitre 3, j'ai étudié les liens bidirectionnels potentiels entre la MOD — à la fois sa concentration et sa composition — et le métabolisme écosystémique dans 548 lacs canadiens. Pour ce faire, j'ai utilisé les taux métaboliques estivaux et les données qualitatives et quantitatives de la MOD compilées dans les chapitres 1 et 2. J'ai réalisé des analyses multivariées et une modélisation par équations structurelles (MES) pour explorer les relations entre les facteurs environnementaux et hydrologiques qui influencent le COD et les taux métaboliques. J'ai ensuite utilisé la MES pour tester les relations spécifiques entre les facteurs environnementaux et hydrologiques qui influencent les pools optiques et moléculaires de la MOD et les différentes composantes du métabolisme. La PPB et la R variaient en fonction des concentrations de TP et de COD. La concentration en COD était associée à un indicateur de l'histoire de l'eau, d'après l'étude du d-excess. La PPB et la R étaient étroitement couplées dans tous les lacs et avaient tendance à converger dans les lacs à forte productivité. La PPB était positivement associée à certaines catégories de la MOD, comme la composante optique C5 et la classe moléculaire aliphatique, et négativement associée aux composés aromatiques et insaturés à haute teneur en oxygène. En revanche, la classe des composés insaturés à haute teneur en oxygène était positivement liée à la R, ce qui suggère que cette classe pourrait alimenter la R. Finalement, je montre que toutes les catégories constituant le pool total de la MOD du lac sont influencées par le métabolisme lacustre, mais que toutes ces catégories de MOD n'ont pas la capacité d'influencer le métabolisme du lac.

Conclusion : Pris ensemble, ces résultats démontrent la forte influence de l'hydrologie sur la concentration et la composition de la MOD, ainsi que ses effets indirects sur le métabolisme du lac. Cette thèse souligne les différences régionales dans le métabolisme de base et la composition de la MOD, différences qui ne sont apparentes que lorsqu'elles sont examinées à une échelle macroscopique (ici, continentale). En démêlant la structure du métabolisme des lacs et la dynamique de la MOD à travers des paysages complexes et à l'échelle continentale, ce travail souligne la nécessité de comprendre l'imbrication de

facteurs multiples à travers de larges échelles géographiques. En outre, il fournit de nouvelles perspectives sur les relations entre des catégories spécifiques de la MOD et les composantes du métabolisme lacustre, faisant progresser ma compréhension du fonctionnement des écosystèmes d'eau douce.

Mots-clés: Métabolisme du lac, matière organique dissoute (MOD), production primaire, connectivité hydrologique, échelle macroscopique, production nette de l'écosystème

ABSTRACT

Lake metabolism and dissolved organic matter (DOM) concentration and composition play crucial roles in shaping ecosystem processes and carbon (C) dynamics in lakes. Despite their importance in biogeochemical cycling, our understanding of their large-scale patterns and regulation has been limited to specific watersheds and small numbers of lakes, often along narrow environmental gradients. Moreover, the potential direct links between lake metabolism and DOM composition are still poorly understood. Here, I conducted a continental-scale study of lake metabolism and DOM—both in terms of concentration and composition—in lakes across Canada as part of the Lake Pulse Network. First, in chapter 1, I aimed to uncover the spatial patterns and environmental drivers of gross primary production (GPP), respiration (R), and net ecosystem production (NEP) across Canadian lakes. Second, I measured various molecular and optical properties of DOM and examined how different molecular and optical properties of DOM relate to in-lake, climate, and hydrological variables (chapter 2). Third, I attempted to disentangle and understand the complex relationships that exist between DOM composition and lake metabolism (chapter 3).

In chapter 1, I examined patterns of lake metabolism across Canada by measuring GPP, R, and NEP during summer. Using an oxygen isotopic ($\delta^{18}O_2$) approach, I assessed metabolic rates in 742 lakes spanning major drainage basins and diverse landscapes. GPP and R varied widely, driven largely by nutrient concentrations, DOC, and chlorophyll. NEP showed weak but significant links to light availability and colored dissolved organic matter (CDOM). My findings reveal regional differences in baseline metabolism and suggest that lakes' sensitivity to eutrophication and browning are shaped by hydrology and climate across the landscape.

In Chapter 2, I explored how different in-lake, catchment, and climate variables influence DOM concentration and composition across lakes in Canada. I analyzed DOC and DOM characteristics in 548 lakes spanning five major drainage basins, using optical measurements, parallel factor analysis, and ultra-high-resolution mass spectrometry. DOM concentration and composition varied both within and among regions, with strong links to hydrology (i.e., deuterium excess (d-excess)) and trophic status (i.e., total phosphorus (TP)). I found region-specific relationships between individual DOM molecular classes and d-excess, highlighting regional differences in DOM dynamics. Although there were differences in reconstructed source-derived DOM composition across regions, hydrologic network processing emerged

as the main driver of regional variation. Overall, my findings highlight that hydrologic processing, more than differences in source inputs, plays a leading role in shaping DOM composition at the continental scale.

In Chapter 3, I investigated potential bidirectional links between DOM—both its concentration and composition—and ecosystem metabolism across 548 lakes in Canada. To do so, I used summer metabolic rates and DOM quality and quantity data compiled in Chapters 1 and 2. First, I applied multivariate analyses and structural equation modeling (SEM) to explore the relationships between environmental and hydrological factors that influence DOC and metabolic rates. I then used SEM to test specific relationships between molecular and optical DOM pools and components of metabolism. GPP and R varied as a function of TP and DOC concentrations. DOC concentration was associated with a proxy of water history, as indicated by d-excess. GPP and R were tightly coupled across lakes and tended to converge in highly productive lakes. GPP was positively associated with certain DOM pools, such as optical components C5 and the aliphatic class, and negatively associated with aromatics and High.O.unsaturated compounds. In contrast, the High.O.unsaturated class was positively related to R, suggesting that this pool may be fueling respiration. Finally, I show that not all the pools within the bulk lake DOM are influenced by lake metabolism, and conversely, not all the DOM pools have the capacity to influence lake metabolism.

Taken together, these findings demonstrate the strong influence of hydrology on both DOM concentration and composition, as well as its indirect effects on lake metabolism. This thesis underscores regional differences in baseline metabolism and DOM composition, which become apparent only when examined at macroscale (e.g., continental) levels. By unraveling the structure of lake metabolism and DOM dynamics across complex continental landscapes, this work emphasizes the need to understand the interplay of drivers operating across regional scales. Furthermore, it provides new insights into the reciprocal relationships between specific DOM pools and metabolic components, advancing my understanding of freshwater ecosystem functioning.

Keywords: Lake metabolism, dissolved organic matter, primary production, hydrological connectivity, macroscale, net ecosystem production

INTRODUCTION

0.1 Context

Freshwater ecosystems are critical links between terrestrial and oceanic environments, forming a continuum that supports the transfer of energy, materials, and biodiversity. Historically viewed as passive transmitters of carbon (C) (Falkowski et al., 2000), they are now recognized as key players in transporting, transforming, and producing C (Cole et al. 2007; Battin et al. 2009). Lakes, although occupying a relatively small area of highly heterogeneous landscapes, play an important role in the C cycle. Lakes mediate the fate of 20% of the world's soil organic carbon (SOC) pool (Sothe et al. 2022), highlighting their disproportionate influence on C dynamics at both regional and global scales. Recent studies have shown that these lakes are sensitive to anthropogenic impacts and climate change (Dupont et al., 2023), which can alter the physicochemical properties of lakes and their C processing capacities and, in turn, affect the broader macroscale C cycle (Schindler et al. 1996; Smol et al. 2005; Marchand et al. 2009; Deemer et al. 2016).

Ecosystem metabolism plays a central role in the transformations that govern C cycling and energy flow, providing critical insights into the fundamental functioning of ecosystems. Gross primary production (GPP), respiration (R), and net ecosystem production (NEP) are the main components of metabolism, which can be measured using several methods. In principle, when gross primary production (GPP) exceeds ecosystem respiration (R), the ecosystem is considered autotrophic (GPP > R, resulting in a positive net ecosystem production, +NEP). In this state, the ecosystem fixes more carbon through photosynthesis than it releases through respiration, thereby storing organic carbon and acting as a net sink of atmospheric CO_2. Autotrophic ecosystems are sustained primarily by internal primary production, with autotrophs (e.g., macrophytes, algae, and photosynthetic microbes) generating the organic matter that supports the food web. Conversely, when the decomposition of organic matter and ecosystem respiration exceed GPP, the ecosystem is heterotrophic (GPP < R, resulting in a negative net ecosystem production, –NEP). Heterotrophic ecosystems rely on the consumption and mineralization of organic matter produced elsewhere or accumulated previously, releasing more CO_2 to the atmosphere than they sequester. Such systems are driven largely by heterotrophic organisms (e.g., bacteria) that decompose organic material, making the ecosystem a net source of CO_2.

In addition, dissolved organic matter (DOM) also is a key factor shaping aquatic environments (del Giorgio and Peters 1993; Solomon et al. 2015; Zwart et al. 2016) and directly influences lake metabolism dynamics (Prairie et al., 2002; Hanson et al. 2003; Solomon et al. 2013). DOM can originate from both autochthonous sources (e.g., phytoplankton, aquatic plants) and allochthonous sources (e.g., inputs from the watershed). Pelagic metabolism and DOM decomposition rates depend on DOM quality and availability through various physical, chemical, and biological processes. Studying metabolic rates and DOM in lakes provides fundamental information on the mechanism underlying lake greenhouse gas (GHG) dynamics, albeit there are a number of obstacles limiting our understanding of patterns and drivers of metabolism and DOM.

There are over 117 million lakes globally (Verpoorter et al. 2014), distributed across heterogeneous landscapes that vary greatly in terms of geology, land use, and hydroclimate. The interplay of multiple natural and anthropogenic drivers results in large variations in metabolic rates within lakes over time and space, and among lakes (Hanson et al. 2008). Lake metabolic pathways are complex, multi-scale processes that cannot be fully understood by examining them through a single lens, because their drivers often transcend the boundaries of scales. For example, regional drivers (e.g., regional land-use) can interact with local drivers (e.g., total phosphorous) (Soranno et al. 2014; Lapierre et al. 2018; Bogard et al. 2020), and such multiscale interactions can lead to widely varying spatial patterns of ecosystem processes. Although the number of aquatic metabolic studies has greatly increased, most work has focused on quantifying community-level rates using in vitro incubation techniques that may not reflect ecosystem-level metabolism (del Giorgio and Peters 1994; Fee et al. 1996; Ask et al. 2012; Seekell et al. 2015).

A major challenge still remains in generalizing findings from studies focused on specific regions (Hanson et al. 2003; Bogard et al. 2019; Bogard et al. 2020, Ayala-Borda et al. 2024) to the millions of lakes distributed across diverse regions. Nonetheless, a few studies have addressed metabolism on broader geographical scales (Yvon-Durocher et al. 2012; Solomon et al. 2013). A number of these studies have found links between metabolic rates and intrinsic lake properties, such as lake size, lake morphometry, nutrient availability, and lake temperature (del Giorgio and Peters 1993; Hanson et al. 2003; Bogard et al. 2020). Faithfull et al. (2011) analyzed published field and experimental studies and concluded that both nitrogen (N) and phosphorous (P) were strong primary and secondary production. In other studies, lake shape also has been shown to be an important predictor of phytoplankton productivity (e.g., Fee et al. 1996).

On the other hand, apart from internal drivers, external drivers also influence lake metabolism. For example, the location of a lake within its catchment, as well as land-use activities in the surrounding watershed, have been found to strongly influence lake metabolism (del Giorgio and Peters 1993; Staehr et al. 2010). Catchment size and characteristics determine the magnitude and type of inputs that reach a lake: larger catchments or those with steep slopes often deliver greater hydrological and nutrient inputs, which can enhance primary production but also increase organic matter loading that fuels respiration. Similarly, land-use practices such as agriculture, urbanization, or forestry alter nutrient and organic matter fluxes. Agricultural catchments, for instance, typically contribute elevated nitrogen and phosphorus, stimulating gross primary production, whereas forested or peat-dominated catchments export high loads of terrestrially derived dissolved organic matter, which can increase heterotrophic respiration and reduce light availability for autotrophs. In addition, hydrological flushing events can deliver pulses of labile DOM, stimulating microbial activity and respiration (Sadro et al. 2012). To unravel how lake metabolism varies across macroscale landscapes, a holistic approach is needed—one that integrates geological, hydrological, climatic, and anthropogenic drivers. Current understanding remains fragmented, particularly regarding interactions between internal (e.g., nutrient, size) and external (e.g., land-use, hydrology) drivers. Taken together, our understanding of how these aquatic ecosystems function under the influence of contemporary natural and anthropogenic pressures remains incomplete, and there is a lack of knowledge on how internal and external drivers interact to shape lake metabolism. Cross-system comparisons of ecosystem-scale metabolic rates in response to a broad array of environmental, land use, and climatic variables have the potential to bridge this knowledge gap.

A second challenge is that DOM concentration and composition can mask the presence of distinct molecular fractions, each potentially shaped by different in-lake and regional drivers. Sobek et al (2007) suggested that lake DOC concentrations in certain regions are controlled by climate and topographic conditions, but that DOC concentration within a given lake is regulated by in-lake properties. Topography can directly influence DOC concentrations through its effects on hydrology, connectivity, and organic matter export. The concentration and composition of DOM in lakes are influenced by various allochthonous and autochthonous sources as well as the biotic and abiotic degradation and transformation of DOM (Kothawala et al. 2015; Ward et al. 2017).

Many studies have found links between DOM concentration, composition and lake internal properties such as water residence time (WRT) (Johnston et al. 2020; Pugh et al. 2021; Kurek et al. 2023; Abbasi et al.

2024). DOM molecular-level composition was also linked to the change of phytoplankton community in lakes over a large range of trophic gradients (Wen et al. 2022). Furthermore, several studies have established connections between DOM composition and concentration and watershed properties such as land cover. For example, Lapierre & del Giorgio (2014) showed that catchments with a high proportion of wetlands export large quantities of DOC that are typically aromatic and terrestrially derived, while agricultural or urbanized catchments deliver more labile, nutrient-enriched DOM. Importantly, DOM inputs—particularly CDOM—are often co-delivered with nutrients, which can exert contrasting effects on primary production depending on concentration and context. At lower DOC concentrations, the nutrient subsidies associated with CDOM inputs, especially phosphorus, can stimulate GPP by alleviating nutrient limitation (Kelly et al. 2018). Zwart et al. (2016) further demonstrated that increases in GPP were linked not only to phosphorus enrichment but also to a shallowing of the upper mixed layer, which improved light conditions for phytoplankton growth. However, at higher DOC concentrations, strong light attenuation by CDOM can suppress GPP despite the presence of nutrients.

Multiple studies have shown that hydroclimatic factors, particularly precipitation and hydrology, are the primary drivers of allochthonous DOM inputs into lakes across diverse regions (Pugh et al. 2021; Kellerman et al. 2020), regulating both DOM production and export within aquatic systems (Fasching et al. 2016; Butturini et al. 2016; Casas-Ruiz et al. 2017). These hydrologically mediated transport processes generate systematic variations in DOM composition, shaping the dominant sources and degradation pathways of organic matter (Lapierre & del Giorgio, 2014; Jones et al. 2016; Massicotte et al. 2017; Hutchins et al. 2017; Cory & Kling 2018; Kellerman et al. 2018). Hydrology, in particular, serves as a key unifying factor linking the diverse variables that influence DOM dynamics in lakes (Jones et al. 2024). Ongoing land-use changes and global warming can alter precipitation and evaporation patterns, thereby modifying hydrological connectivity. As a result, some lakes may become more connected to their surrounding watersheds and surface runoff, while others may become increasingly isolated. These shifts can influence both the concentration and composition of DOM in aquatic ecosystems. However, the extent and strength of these changes at a continental scale remain largely unexplored.

Although there have been some exceptions (Kellerman et al. 2014; Lapierre et al. 2015; Johnston et al. 2020; Pugh et al. 2021; Kurek et al. 2023), most studies have focused on patterns and drivers of DOM concentration and composition within specific catchment or narrow geographic scales using small sets of predictors and/or response variables. Therefore, the specific climatic and hydrologic drivers that regulate

overall lake DOM concentration and composition at the continental scale under current and future climates and land use conditions remain unclear. Patterns and drivers of DOM concentration and composition at this scale might be the result of multiple drivers operating at various local and regional scales (Soranno et al. 2014). Therefore, it is uncertain whether the drivers identified at the watershed or local regional scale also operate across larger geographic scales, for example, across a continental gradient, or whether these drivers are important factors in different hydroclimates.

A third major challenge is the complex nature of the relationship between metabolism, DOM concentration and composition. DOM dynamics are shaped by lake metabolism through processes such as algal production of organic carbon, selective removal of specific DOM fractions, and microbial transformation and molecular restructuring of the DOM pool (Stadler et al. 2020). Terrestrially derived colored DOM may have a negative effect on GPP by altering underwater light regimes and reducing thermocline depth (del Giorgio et al. 1994; Karlsson et al. 2009), but DOM also can enhance R by providing C sources for bacteria and indirectly for other heterotrophic organisms (e.g., Hessen 1992). Therefore, both autochthonous DOM and allochthonous DOM are linked to lake metabolism but in very different ways (del Giorgio et al. 1999; Hanson et al. 2003; Pace and Prairie 2005; Solomon et al. 2013). Furthermore, an increasing number of studies suggest that global changes such as acidification recovery, climate, and land use changes result in shifts in the amount and composition of DOM exported to lakes (Winterdahl et al. 2014). Even though several studies have shown the link between lake metabolism and the quality and quantity of DOM (Ayala-Borda et al. 2024), we still lack a comprehensive understanding of which specific aspects of metabolism such as GPP, R and NEP are most strongly influenced by variations in DOM composition and concentration. Similarly, there is limited understanding of how different DOM pools (e.g., based on fluorescent indices and molecular properties) directly affect metabolism components. Addressing this knowledge gap would provide insights into ecosystem-scale C fluxes, linking terrestrial inputs, aquatic processing, which are critical for understanding the global C cycle.
Together, these issues underscore major gaps in our knowledge of lake functioning and its significance in regulating the aquatic carbon cycle.

0.2 Thesis objectives

To bridge these knowledge gaps, this thesis aims to understand the complex interaction between local and regional variables that determine lake metabolism and DOM dynamics across large spatial and environmental gradients (figure 0.3). This goal is addressed through three sequential objectives:

1. The first main goal of this thesis is to understand large-scale patterns in lake metabolic rates (i.e., GPP, R and NEP) across wide climatic, landscape and environmental gradients, and determine how drivers at different scales interact to control lake ecosystem metabolism at macroscale across Canadian lakes (figure 0.1).

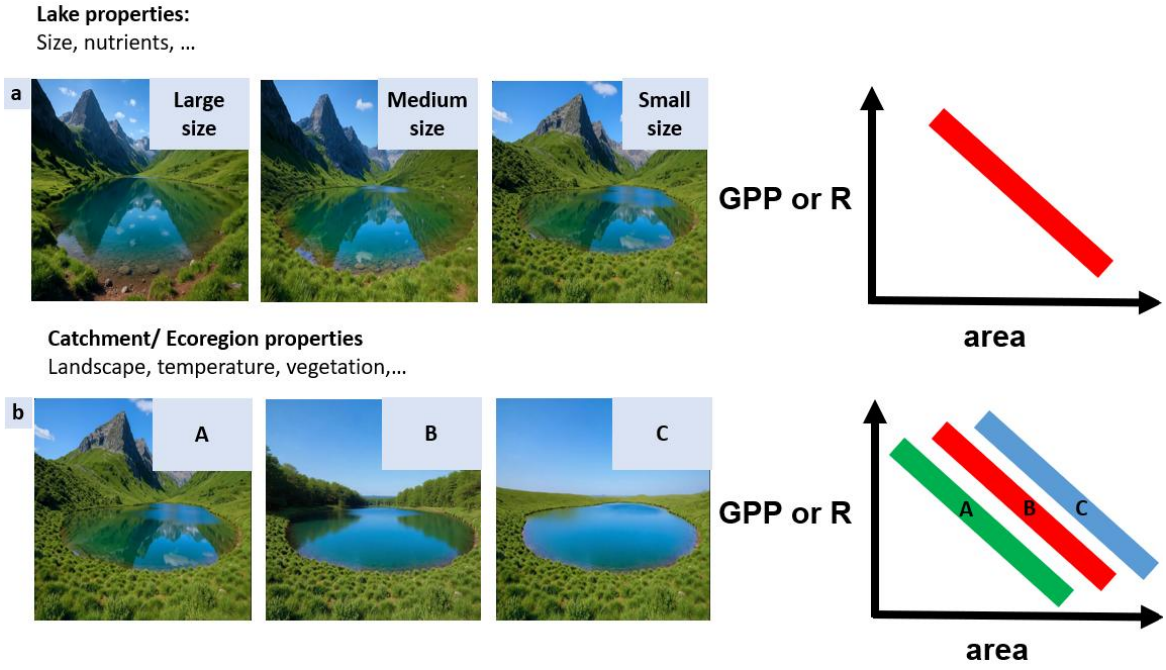

Figure 0.1: Hypothetical illustration of how both in-lake and catchment/ecoregion properties influence lake metabolic rates (GPP or R). (a) Shows lakes of varying size within the same region, where a negative relationship is observed—larger lakes exhibit lower GPP or R. (b) Lakes with similar size situated across different ecoregions, each showing distinct scaling relationships between GPP or R and area. This variation highlights the role of catchment and ecoregion characteristics, such as landscape structure, temperature, and vegetation, in modulating lake metabolism. Together, the panels emphasize the importance of both local (in-lake) and regional (ecoregional) drivers in shaping macroscale patterns of ecosystem metabolism across Canadian lakes.

2. The second major objective is to investigate large-scale patterns in the concentration and composition of DOM in lakes across Canada and identify the key drivers of DOM concentration and composition at large-scale. Specifically, we aimed to examine how broad hydrological gradients, inferred from water isotopes, modulate the effects of local and regional factors on continental-scale variation in lake DOM quantity and quality (figure 0.2).

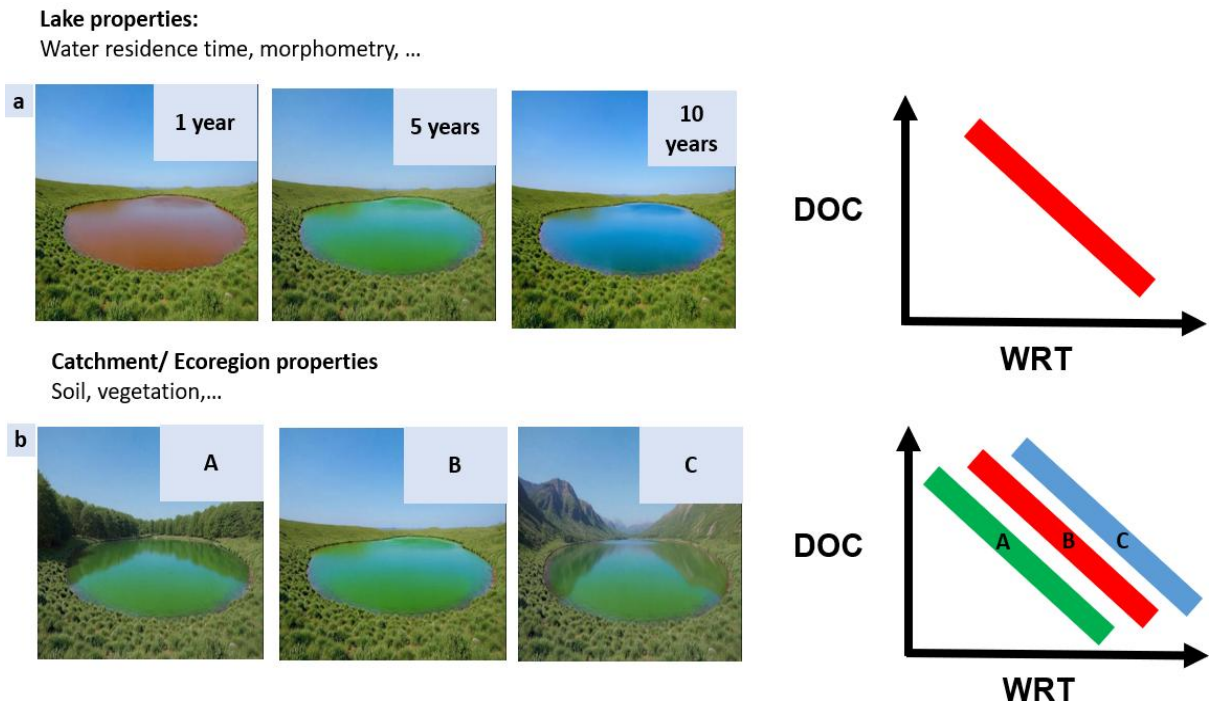

Figure 0.2: Hypothetical illustration of how both in-lake and catchment/ecoregion properties influence lake DOC concentration. (a) Shows lakes of varying WRT within the same region, where a negative relationship is observed— lakes with higher DOC tend to have lower DOC. (b) Lakes with similar WRT situated across different ecoregions, each showing distinct scaling relationships between DOC and WRT. This variation highlights the role of catchment and ecoregion characteristics, such as soil and vegetation, in modulating DOC. Together, the panels emphasize the importance of both local (in-lake) and regional (ecoregional) drivers in shaping macroscale patterns of DOC across Canadian lakes.

3. The third major objective is to understand the relationship between different optical and molecular DOM pools and lake metabolism (GPP and R). We aim to better understand how DOM quantity and quality mediate lake metabolism, and to identify optical and chemical components of DOM that either influence or are produced by GPP, and what components of DOM influence R or are influenced by R.

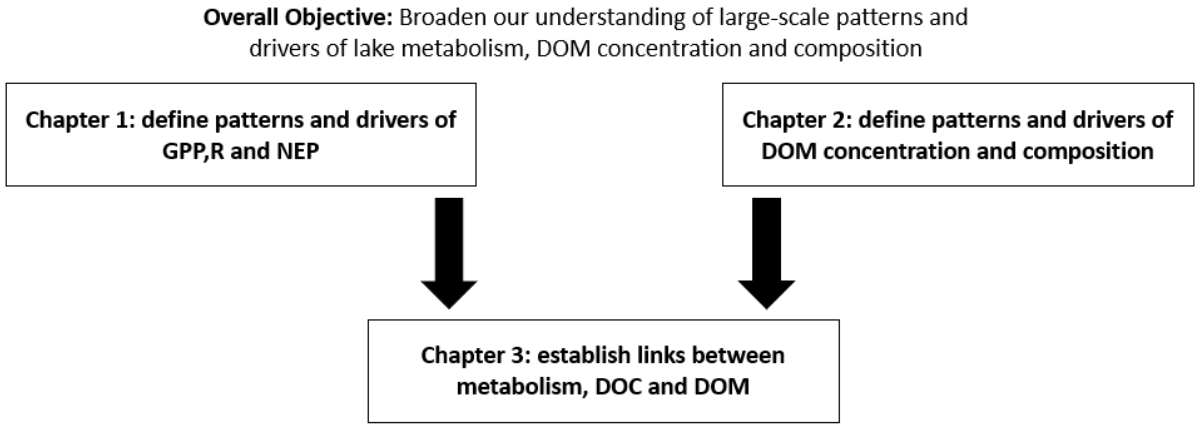

Figure 0.3: Overall thesis objectives and chapters specific goals.

0.3 General approach

0.3.1 Measuring lake metabolism

In recent decades, several methods and techniques have been developed and used to quantify GPP, R, and NEP within aquatic ecosystems. The suite of direct and indirect methods of aquatic ecosystem metabolism each have their strengths and weaknesses. Incubation of water in bottles under both light and darkness has been widely used to determine aquatic ecosystem metabolism. Traditionally, this method has been used to obtain planktonic community photosynthesis using both O_2 and ^{14}C-bicarbonate tracer additions, with incubations spanning time scales from hours to days (Kemp and Boynton 1980; Smith and Kemp 1995; Gazeau et al. 2005).

These incubation-based approaches have advantages and disadvantages. Incubation-based studies can distinguish metabolic components (GPP, R and NEP) and when bottle incubation is combined with sediment chamber incubation could help to differentiate the contribution of benthic and pelagic processes to the ecosystem-scale metabolism (Gazeau et al. 2005). The rates of production and respiration derived from bottle incubations were considered to give ambiguous results or even fail to integrate large spatial and temporal scales (Bender et al. 1987). Today, bottle incubations are considered to be artificial systems that introduce many container-associated artifacts such as excluding the grazer community or preventing mixing processes in the water column (Staehr et al. 2010).

Measurements of diel changes in ambient DO concentration have been used to directly determine GPP and R. Such open water techniques were used in a variety of aquatic ecosystems over the past decades to avoid the limitations of bottle and chamber incubations (e.g., Sargent and Austin 1949; Kemp and Boynton 1980; Barnes 1983; Gattuso et al. 1993; Caffrey 2004; Staehr and Sand-Jensen 2007; Bogard and del Giorgio 2016). As an *in-situ* method, it derives GPP and R by tracking the inputs and outputs of photosynthesis and respiration (O_2 or CO_2). For the accuracy of these methods, it is critical to derive reliable estimates for the exchange rates of O_2 with the atmosphere, yet their high uncertainty in time and space can be the method's major limitation. Additionally, metabolism estimates often require the conversion of rates in oxygen and carbon units, and the uncertainties in the O_2:C conversion factor may be substantial (Staehr et al. 2010). Despite these limitations, diel O_2 or CO_2 methods are a powerful replacement for bottle and chamber methods since they clearly avoid the many container-associated artifacts. Also, diel changes in DO concentration measure all system components, enabling ecosystem-

level metabolic estimates. Apart from all the advantages, using these probes across multiple systems has many logistical challenges. Another method to estimate metabolism in aquatic systems is to use a combination of ambient oxygen concentration, and oxygen isotopic composition ($\delta^{18}O$) (Bogard et al, 2017). The Isotopic approach has been used to estimate metabolism using the fact that oxygen naturally has three stable isotopes (^{16}O, ^{17}O and ^{18}O) of varying atomic abundance. In the atmosphere, the ratio of stable isotopes initially depends on the isotopic composition of oxygen which has been produced photosynthetically from vegetation in both terrestrial and aquatic ecosystems, and also depends on isotopic fractionation during O_2 removal by respiration processes (Staehr et al. 2010). The oxygen isotopic approach provides an integrative snapshot of mixed layer metabolism in the case of stratified lakes, or whole lake metabolism in the case of shallow lakes (Bogard et al. 2017) which also can obtain episodic production events. There are some disadvantages related to the oxygen isotopic method. First, it requires simultaneous estimation of air-water exchange rates, which are highly variable temporally and spatially. Second, it is not possible to distinguish abiotic and biotic fluxes of dissolved oxygen, which potentially artificially increase measured rates of respiration including all the individual pathways (Bogard et al. 2017). Approaches requiring intensive in-situ and lab-based methods, such as incubations or the deployment and maintenance of sensors, are challenging and often impractical for remote lakes or large-scale studies across many lakes. In this thesis, I used the oxygen isotope method, as it is uniquely suited for studying metabolism across many systems, providing a practical and reliable approach for large-scale investigations.

0.3.2 Qualitive measurements of DOM

Over the past six decades, measuring DOC concentrations has been a standard practice in freshwater and biogeochemical studies. DOM is a complex and heterogeneous mixture of compounds derived from both autochthonous (within-lake) and allochthonous (catchment-derived) sources. This complexity makes it difficult to resolve the individual components of DOM. While bulk DOM measurements can offer valuable information, they provide limited insight into the diversity and composition of its constituent compounds. DOM composition can be studied with multiple approaches, each enabling us to disentangle this complex organic mixture.

Optical and absorbance measurements of DOM provide insights into its sources and molecular weight, structure, and degradation of DOM (Stedmon et al. 2003). Absorbance in the UV range, commonly reported as specific ultraviolet absorbance (SUVA), is useful in estimating aromaticity associated with terrestrial DOM (Weishaar et al. 2003). Chromophoric dissolved organic matter (CDOM) is the fraction of

DOM that absorbs ultraviolet and visible light (Stubbins et al. 2008), while fluorescent dissolved organic matter (FDOM) represents a subfraction of CDOM (Coble et al. 1990). Fluorescence excitation-emission matrix (EEM) spectroscopy, combined with statistical approaches such as parallel factor analysis (PARAFAC), offers valuable information to distinguish between autochthonous FDOM (e.g., derived from algal, bacterial production, and macrophytes) and allochthonous FDOM (e.g., terrestrial sources) (Stedmon et al. 2003).

Stable isotopes (^{13}C and ^{12}C) and radioactive isotopes (^{14}C) have been used to provide valuable insights into the relative contributions of DOM sources and trace carbon pathways within the food web (Grey et al. 2001). High-performance liquid chromatography–size exclusion chromatography (HPLC–SEC) is a widely used method for investigating the molecular weight distribution of DOM in aquatic ecosystems. This technique distinguishes between small and large molecules, aiding in the identification of DOM sources (Yan et al. 2012). However, SEC is a separation method and requires coupling with an appropriate detector to characterize the eluted fractions.

Nuclear magnetic resonance (NMR) spectroscopy is another method widely used to reveal the types and proportions of chemical bonds and functional groups present in DOM molecules. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) is a state-of-the-art technique that has been used to examine the molecular properties of DOM, further enhancing our understanding of its source, chemical composition, structure, and elemental composition (Minor et al. 2014; Nebbioso and Piccolo 2013). FT-ICR MS has been widely employed to characterize DOM sources across a range of aquatic ecosystems, including rivers (Spencer et al. 2019), groundwater (McDonough et al., 2020), and lakes (Johnston et al., 2019; Kellerman et al. 2015). When combined with other analyses, FT-ICR MS provides a robust approach for identifying DOM origins and elucidating the key processes governing its composition and transformation (Behnke et al. 2021; McDonough et al. 2020; Stubbins et al. 2017). In my thesis, I studied DOM composition using two different approaches that consider both optical and molecular properties of lake DOM. I used fluorescence spectroscopy to construct EEMs and utilized the advanced statistical tool PARAFAC, to identify the main optical components of the DOM. Additionally, I used FT-ICR MS to analyze the molecular composition of DOM, providing detailed insights into its chemical composition.

0.3.3 Lake sampling

Most of the data used in this thesis are from the LakePulse network. The LakePulse Network is the first pan-Canadian initiative to comprehensively study lakes, covering a wide diversity of lake types along large gradients of catchment and climate features across various ecoregions in Canada. This network has been documented in detail previously (Huot et al., 2019). From over a million lakes in Canada, a subset of lakes was selected in 12 ecoregions out of the 16 Canadian ecoregions so that it adequately represents lakes of different sizes (≥0.1 to 0.5 km^2, ≥0.5 to 5 km^2, ≥5 to ≤100 km^2) and human impact (low, medium, high). Regions with similar ecological, geological, and climatic properties are defined as ecoregions. Furthermore, maximum distance of 1 km from a road also was among the selection criteria for accessibility reasons. This project provided an excellent opportunity to discriminate the relative role of lake-specific (i.e. size, shape, nutrient, etc.) versus eco-regional (i.e. vegetation type, land use, etc.) drivers of lake metabolism and DOM. I also incorporated an additional subset of lakes (n = 78), sampled within the framework of the Landscape Carbon Accumulation through Reduction in Emissions (L-CARE) program, which was specifically used in the first chapter of the thesis.

0.3.4 Field measurements and derived metrics

All selected lakes were sampled once between 2017 and 2019 using a 3-meter (10-foot) aluminum boat. During each visit, we collected a comprehensive set of limnological parameters following standardized protocols. To reduce the variability associated with seasonal cycles, sampling was carried out during the period of maximum thermal stratification (July to the beginning of September). It should be noted that by focusing on the summer period, this study provides a snapshot that may not represent lake conditions during other critical seasons, such as spring mixing or winter under-ice cover. Furthermore, to minimize diel variability in lake characteristics sampling was typically conducted from mid-morning to early afternoon. Epilimnetic samples were taken at the deepest point of the lake. When bathymetry was not available, briefly, we used a motorboat to cross the lake or a specific bay in the lake while continuously monitoring the depth. To quantify the human impact index on each lake, land use in each watershed was evaluated. Land-use patterns were obtained from existing 30-meter resolution remote sensing data (discussed in detail in Huot et al. 2019).

Meteorological parameters (air temperature, relative humidity, wind speed, atmospheric pressure, precipitation, and cloud cover) were recorded in situ at each sampling location using a Kestrel 5500

portable weather station. To complement field measurements, we obtained historical hourly wind data for the one-week sampling period from the nearest Environment Canada meteorological station. For the 30-day period solar radiation and precipitation data, we used ERA5-Land hourly data (1981-present; available at: https://doi.org/10.24381/cds.e2161bac), which were processed using GIS-based methods detailed in Huot et al. (2019).

Water temperature, dissolved oxygen, pH, salinity, conductivity, and chlorophyll *a* vertical profiles were taken through the water column using an RBR Maestro multi-parameter water quality meter (RBR Ltd. Ottawa, Canada). Epilimnetic water samples were collected using an acid-washed 2-m integrated tube sampler and immediately transported to a mobile shore-based laboratory for processing. Samples were preserved according to analytical requirements, with storage in portable -80°C freezers, -20°C freezers, refrigerated conditions (4°C), or at ambient temperature. All samples were shipped to their respective analytical laboratories on a weekly basis. Samples for DOC and DIC, TN (mg L^{-1}), and TP (µg L^{-1}) were analyzed at the GRIL laboratory, located at the Université du Québec à Montréal.

To estimate water column light penetration, daily means of surface photosynthetically active radiation (PAR) for the 15 days before sampling were obtained from CERES_SYN1deg_Ed4A (Wielicki et al. 1996) measurements. The ASTM G173-03 surface reference spectrum (a reference irradiance spectrum of direct sunlight on Earth's surface at the ground level) (ASTM 2003) was used as the direct irradiance spectrum (incident irradiance only from the direction of the Sun), while the diffuse spectrum (incident irradiance from angles different from the direction of the direct sunlight, scattered by the atmosphere and clouds) was adjusted following Bartlett et al. (1998) with a cloud cover factor of 1. Day lengths and solar zenith angles (the angle between the local vertical (zenith) and the line connecting the observer to the Sun) were obtained from the R package *Suncalc*. The mean PAR was corrected for daytime only before averaging. Backscattering coefficient (bb) was measured for 73 lakes using HydroScat-6 (HOBI Labs, United States). The remainder was estimated using a boosted regression tree model that incorporated physicochemical and watershed variables, including *Chla*, TSS, and Secchi depth. The attenuation coefficient (k_d) was calculated following (Lee et al. 2005) and validated with in situ measurements obtained with the Hyperpro (Satlantic, Canada) and C-OPS (Biospherical Instruments, United States) profiling probes. Average epilimnion irradiance (E_{wc}) was calculated as the result of the integral calculation:

$$\widehat{E_{wc}} = \frac{\int_{Z_{mix}}^{z=0} E(z)\, dz}{Z_{mix}} \qquad (1)$$

which, after separately adding the direct and diffuse irradiances results in:

$$\widehat{E_{wc}} = \frac{E_{Dir}(1-e^{-k_d^{Dir} Z_{mix}})}{k_d^{Dir} Z_{mix}} + \frac{E_{Dif}(1-e^{-k_d^{Dif} Z_{mix}})}{k_d^{Dif} Z_{mix}} \quad (2)$$

Where E_{Dir} and E_{Dif} represent the direct and diffuse surface irradiance, k_d^{Dir} and k_d^{Dif} denote the direct and diffuse k_d values (calculated differently due to distinct incidence angles), and Z_{mix} is the mixing depth. We estimated the mixed layer depth (Z_{mix}) from temperature profiles and water density gradients (threshold 0.1 kg m^{-3} per meter) in cases of thermal stratification, we used the "Lake Parameters" add-in in JPM software (version Pro 16.0.0).

We obtained the data for WRT from the HydroLakes database (Messager et al., 2016) and to estimate the movement of water between land, sea, and atmosphere and have an integrative perspective of the hydrologic history, we employed an isotopic approach measuring oxygen ($\delta^{18}O$) and hydrogen ($\delta^{2}H$) stable isotopes. Water isotopes were measured using a LGR DT-100 Liquid Water Stable Isotope Analyzer (Los Gatos Research Inc., Mountain View, CA) at the GEOTOP laboratory, located at the Université du Québec à Montréal. Autosampler vials were filled with 1 ml of sample and 6 injections per sample were analyzed. The measured values were post-processed for quality control (based on stability of injection volume, pressure and variation between injections). The first two injections were always rejected and the average of the remaining measurements that passed the quality check was used as the sample value. The isotopic composition was calculated as:

$$\delta^{18}O \text{ or } \delta^{2}H = ((R_{sample} - R_{standard}) / R_{standard}) \times 1000 \quad (3)$$

where R_{sample} and $R_{standard}$ represent the isotope ratios ($^{18}O/^{16}O$ or $^{2}H/^{1}H$) of the sample and the Vienna Standard Mean Ocean Water (VSMOW), respectively. Deuterium excess (*d-excess*; Dansgaard, 1964) was derived as:

$$\text{d-excess} = \delta^{2}H - 8 \times \delta^{18}O \quad (4)$$

This metric reflects evaporation dynamics: higher *d-excess* indicates less evaporated (recent) precipitation, while lower values suggest evaporative enrichment (Turner et al., 2014).

CHAPITRE 1

The interaction of regional and local drivers shapes summer ecosystem metabolism in lakes across Canada

Amir Reza Shahabinia[1,2], Matthew J. Bogard [3], Paul A. del Girogio[1,2]

[1] Department of Biological Sciences, University of Quebec at Montreal, Montreal, QC, Canada

[2] Interuniversity Research Group in Limnology (Groupe de Recherche Interuniversitaire en Limnologie; GRIL), Montreal, Quebec, Canada

[3] Department of Biological Sciences, University of Lethbridge, Lethbridge, AB, Canada

Published at Limnology & Oceanography (2025) DOI: 10.1002/lno.70095

Key Words: Primary production; Metabolism; Respiration; Net ecosystem production; Food web; land use; Climate; Catchment

N.B. References cited in this chapter are presented at the end of the thesis

1.1 Abstract

Assessments of lake gross primary production (GPP) and respiration (R) and their balance (net ecosystem production, NEP) have been limited to specific watersheds and limited number of lakes, often along narrow environmental gradients. This is because conventional approaches require either lengthy incubations or the deployment of monitoring equipment, none of which are feasible for large-scale studies. Here we present a macro scale study of lake metabolism and explore the patterns and drivers of GPP, R, and NEP in lakes across Canada as part of the LakePulse network. We measured summertime water column metabolic rates in 742 lakes, using an oxygen isotopic ($\delta^{18}O_2$) approach, which providing an integrative snapshot of mixed-layer metabolism in stratified lakes, or whole-lake metabolism in polymictic lakes. The lakes were distributed across the five major Canadian continental drainage basins, covering a wide range of in-lake, land use, and climatic features. GPP and R varied by four orders of magnitude across lakes and regions, driven by factors such as total phosphorus and nitrogen (TP, TN), dissolved organic carbon (DOC), and chlorophyll. NEP had a weak but significant positive linear relationship with water column light and a negative relationship with colored dissolved organic matter (CDOM). Our results reveal systematic differences in regional baseline GPP and R driven by landscape properties such as altitude, and that lake metabolism in some regions may be more sensitive to eutrophication and browning, mediated by regional hydrology, which is itself linked to climate.

1.2 Introduction

Quantifying and understanding the role of natural ecosystems as C sources and sinks has been the focus of extensive research over the past two decades, especially in the context of rapid climate change. In particular, there is a need to understand what controls this function, and also to predict how this key ecosystem function may shift in the future. In the case of lakes, this involves understanding the controls of CO_2 exchange with the atmosphere and also C burial in lake sediments. In lakes, CO_2 exchange with the atmosphere and C burial are not always coupled, both processes being influenced by biological, physical, chemical, and environmental factors not necessarily in the same combination. Lake metabolism plays a central role in shaping both CO_2 dynamics and organic C burial, and therefore the function of lakes as net C sinks or sources of C to the atmosphere. Understanding the magnitude of gross primary production (GPP), respiration (R) and especially net ecosystem production (NEP = GPP-R), is therefore key to our capacity to understand the current and future role of lakes in the landscape C balance. Lake metabolism is particularly vulnerable to climate and human impacts (Nürnberg et al. 1998; Staehr et al. 2012), and it is therefore important to understand the drivers of lake metabolism to predict how lake C source and sink functioning may vary in response to changing environmental conditions.

Studies of aquatic metabolism, including GPP, R and the net balance between the two, provide valuable insights into fundamental aspects of lake function, such as food web dynamics and the flow of energy and materials within the lake (Staehr et al. 2012). This also serves as an indicator of the relative importance of allochthonous versus autochthonous sources of organic matter, and whether the ecosystems are net autotrophic or heterotrophic and their contributions to broad-scale biogeochemical cycling (del Giorgio and Peters 1993; Zwart and Brighenti 2021). Over the past few decades, the number of studies on aquatic metabolism has significantly increased, and there is a relatively strong understanding of the drivers of lake metabolism. Several studies have established connections between lake metabolism and intrinsic lake properties, such as nutrients, DOC concentrations, light, and temperature (del Giorgio et al. 1999; Prairie et al. 2002; Solomon et al. 2013; Bogard et al. 2020; Holgerson et al. 2022). Likewise, watershed properties, such as land use and local hydroclimate, have also been shown to play a significant role in regulating lake metabolism (Staehr et al. 2010; Yvon-Durocher et al. 2010; Scordo et al. 2022). In this regard, extreme rain events have been shown to replace autochthonous DOC with allochthonous DOC and change autotrophic status to strongly heterotrophic (Sadro et al. 2012). Regional wildfire smoke has been shown to influence lake metabolism differently across gradient of trophic gradient (Smits et al. 2024). Collectively, these

factors may interact to determine whether lakes are net heterotrophic (−NEP) or autotrophic (+NEP) (Corman et al. 2023).

Despite this recent progress, most studies have attempted to understand either metabolism of individual lakes, or identify patterns and drivers of metabolism across lakes within a given watershed or region, using either in-vitro incubation techniques (del Giorgio and Peters 1993; Fee et al. 1996; Ask et al. 2012; Seekell et al. 2015) or more recently free-water approaches, which better reflect ecosystem-level processes (Bogard et al. 2020) most of these comprise a limited number of systems within a narrow geographic scope, although there have been some exceptions (Solomon et al., 2013; Puts et al. 2023; Smits et al. 2024). Across much larger spatial, environmental and climatic gradients, typically at continental scales, there are large variations in metabolic rates among lakes (Smith et al. 1991; Smith et al. 1995; Hanson et al. 2008), and this continental scale variability in lake metabolism has neither been well described nor its underlying drivers well understood. The continental scale patterns in lake metabolism are not a simple extension of the patterns found within any given region or watershed, but rather are likely the result of the interplay of multiple drivers operating at very different spatial and temporal scales, and these complex interactions cannot be fully understood by examining single watersheds or regions, or narrow environmental gradients or lake types (Soranno et al. 2014; Lapierre et al. 2018). Large scale sampling programs (e.g., LakePulse (Huot et al. 2019), and U.S. National Lake Assessment (NLA) (USEPA 2009) must be leveraged to understand how lake metabolism is structured across the complex landscape mosaic that exists at a continental scale.

Here we present a macro scale assessment of lake metabolism across the Canadian landscape, where we seek to determine cross-regional patterns in lake gross primary production (GPP), ecosystem respiration, and the balance between the two (NEP), and to identify local to regional metabolic drivers and their potential interactions underlying these patterns. We assessed summertime GPP, R, and their balance (NEP) in 580 lakes using the oxygen isotopic ($\delta^{18}O_2$) free-water approach that integrates ambient O_2 concentration and isotopic signature, providing a comprehensive perspective of mixed-layer metabolism in the case of stratified lakes, or whole lake metabolism in the case of shallow lakes. The lakes varied greatly in morphometry and physicochemical properties and were located across 11 ecoregions within the 6 major continental basins and along extremely wide climatic, geographic and human impact gradients. This allowed us to address three key questions: 1) What is the variability in metabolic rates in lakes within and across the major continental drainage basins in Canada? 2) Across this wide spectrum of topographic,

environmental, land use, and climatic gradients, what are the primary drivers of metabolic rates across lakes? 3) How do internal (in-lake) and external (region-specific) drivers interact to modulate ecosystem metabolism at the continental scale? This study contributes to our understanding of the magnitude and variability of lake metabolic rates at the continental scale in the Northern Hemisphere, which identifies previously undetected interactions between local and regional features driving lake metabolism at the macroscale.

1.3 Methods

1.3.1 Study sites

To investigate the patterns and drivers of GPP, R, NEP, and GPP:R in Canadian lakes, a comprehensive set of limnological parameters was collected once between 2017 and 2019, from late June to mid-September. These data were obtained from 664 lakes across Canada as part of the Canadian LakePulse Network (Huot et al. 2019), funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), and these lakes were distributed along the entire continental territory of Canada (Figure 1.1). The lakes that we sampled covered the six major ocean drainage areas (continental basins) of Canada: Atlantic Ocean, Great Lakes–St. Lawrence, Gulf of Mexico, Hudson Bay, Arctic Ocean and Pacific Ocean. To balance continental basin observational group sizes, we merged lakes located in Gulf of Mexico (n = 2) to Hudson Bay basin, and therefore we had only five major basins in the study. Continental watershed delineations were shown to best align for biogeography based on an analysis of the spatial patterns in plankton community composition and diversity across the 624 lakes (Paquette et al., 2021). Stratified random sampling design was used to capture the diverse range of lake size and human impacts and ecozone across Canada. The target lakes were randomly selected to be within a 1-km radius of roads because of accessibility constrains. To determine the human impact index for each lake, land use within each watershed was assessed. Land-use patterns were derived from existing remote sensing datasets with a 30-meter resolution, as detailed in Huot et al. (2019). Although our ecozone-based stratification aimed to capture climatic variation, the concentration of sampling in southern ecozones may limit our ability to fully represent Canada's broader hydroclimatic spectrum, particularly for northern lakes. An additional subset of lakes (n = 78) was sampled within the context of the Landscape Carbon Accumulation through Reduction in Emissions (L-CARE) program, managed at the Vale Living with Lakes Centre at Laurentian University, and these lakes were located in the Sudbury region of northern Ontario. In all cases, lakes were visited once to sample a range of limnological parameters within a single day, and one lake was sampled per day during the field campaigns.

To minimize the variability associated with seasonal cycles, sampling was conducted during the period of maximum thermal stratification (from July to early September). Additionally, sampling was typically carried out from mid-morning to early afternoon to minimize diel variability in lake characteristics and to minimize error in metabolic estimates (Bogard et al. 2020). The synoptic sampling design involving a single point sample carried out during a 2-month summer window represents only a snapshot in time, and we are therefore unable to reconstruct seasonal or long-term temporal patterns and drivers (For further detail on Representativeness of summer sampling see Supporting Information). While the primary focus of this study was to investigate large-scale patterns in metabolic rates across lakes at a continental scale, and thus the sampling scheme focused on spatial rather than temporal variability, a subset of lakes (n = 8) was sampled in May, July, and September (2018 and 2019) to assess temporal variability in metabolism (see supplemental material).

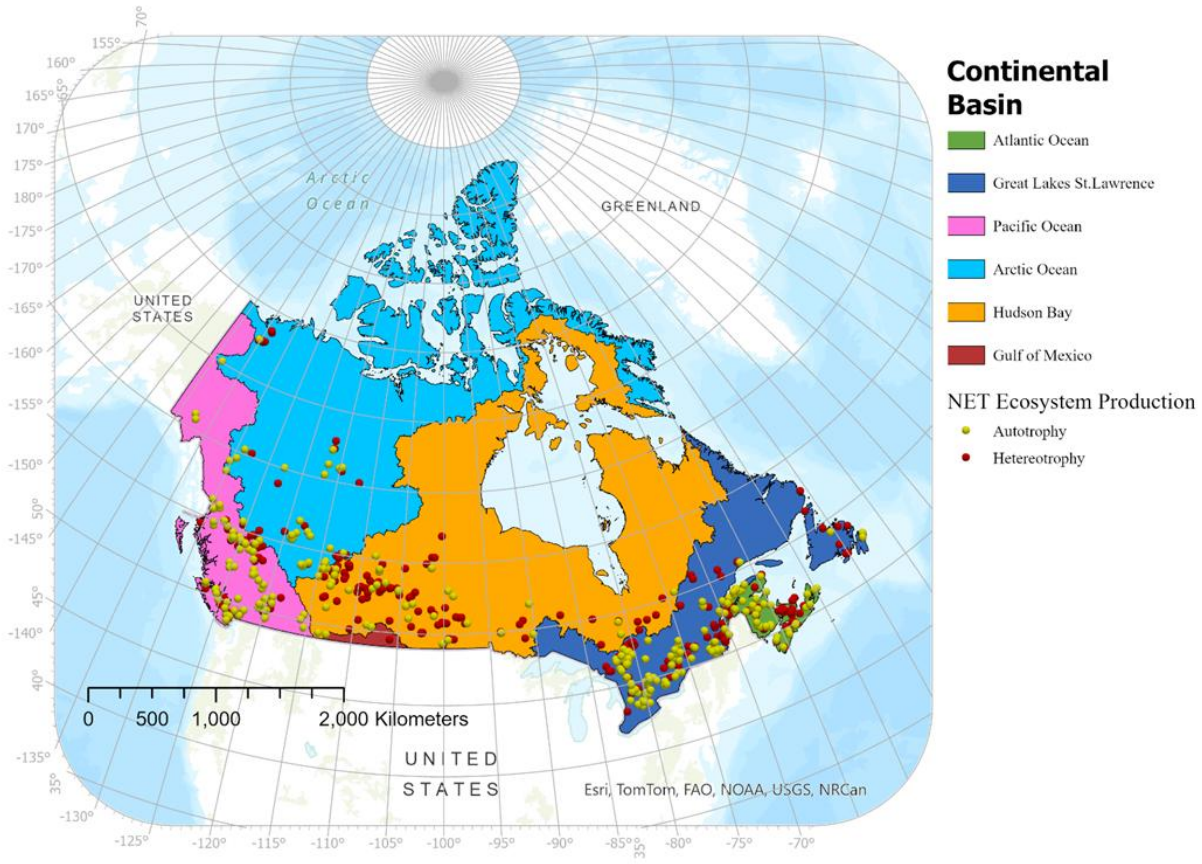

Figure 1.1: Map of sampled lakes by the NSERC Canadian LakePulse Network in Atlantic Ocean, Great lakes St-Lawrence, Pacific Ocean, Hudson Bay, Arctic Ocean continental basins across Canada. Dots with yellow color indicate the Autotrophic and dots with red color shows lakes with Heterotrophic status.

1.3.2 Field sampling

Epilimnetic samples were obtained at the deepest point of each lake using an acid-washed 2-meter tube sampler and transferred to a mobile lab located on the lakeshore for further processing. A summary of the limnological and catchment properties used in this study is presented in Table 1.1. Water temperature, dissolved oxygen, pH, salinity, conductivity, and chlorophyll-a vertical profiles were collected throughout the water column using an RBR Maestro multi-parameter water quality meter (RBR Ltd., Ottawa, Canada). Meteorological parameters, including air temperature, relative humidity, wind speed, atmospheric pressure, precipitation, and sky conditions (cloudiness), were measured at each site using a Kestrel 5500 weather meter, yellow (field protocols described in NSERC Canadian LakePulse Network 2021). Additionally, hourly wind data during the sampling period (7 days prior to sampling), precipitation, and solar radiation (30 days prior to sampling) were obtained from the nearest meteorological station in each region. We also obtained lake morphometric variables from HydroLAKES v. 1.0 (Messager et al. 2016). Triplicate samples for $\delta^{18}O_2$ analysis were collected from the surface waters at the deepest point in each lake and preserved in 12 mL vials. These vials had been pre-evacuated and flushed with He gas to eliminate traces of O_2 and CO_2, and they were preserved with 120 µl of saturated $ZnCl_2$ solution. For isotopes of hydrogen (δ^2H) and oxygen $\delta^{18}O_2$ in water, samples (H_2O) were taken from the same point in the surface water of each lake, stored in 30 mL HDPE bottles without any air bubbles, and subsequently kept at 4℃ in the dark until further analyses.

Table 1.1:Limnological and catchment characteristics of the lakes used in this study.

Parameter	Abbreviation	Unit	Median	Mean	St.dev	Range
Maximum Depth	max_depth	m	6.6	9.9	10	1 - 64.1
Lake area	area	Km^2	0.6	5.2	16.3	0.007 - 266
Chlorophyll a	chla	μgL^{-1}	2.9	6.8	10.7	0.04 - 72
Total phosphorus	TP	μgL^{-1}	18.8	29.2	23.9	2 - 99
Total nitrogen	TN	mgL^{-1}	0.2	0.58	0.8	0.01 - 11.9
Dissolved organic carbon	DOC	mgL^{-1}	8.2	13	12	0.20 - 220
Dissolved inorganic carbon	DIC	mgL^{-1}	13.4	28.3	47.8	0.05 - 432
Water temperature	water_ temperature	C	21	20.6	2.5	13 - 28
CDOM 440	CDOM	m^{-1}	1.3	1.8	1.6	0.006 - 9.7
pH	pH	n.a	8.2	7.4	2.3	0.002 - 9.9
Catchment area	catchment	Km^2	17.9	288	1775	0.2 - 37459
Water residence time	WRT	day	464	1070	1579	0.1 - 9365
deuterium excess	D-excess	‰	-2.3	-6.24	12.3	-37.3 - 21.1
Secchi depth	secchi	m	2.4	2.9	2.2	0.07 - 15
Epilimnion irradiance	water_light	umol photo m^{-2} s^{-1}	241	247.9	135	7.89 – 777
Fraction of agriculture	F_agriculture	%	0	0.094	0.19	0 - 0.86
Fraction of urban	F_urban	%	0.018	0.068	0.13	0 - 0.88
Fraction of forest	F-forest	%	0.0014	0.019	0.04	0 - 0.49
Fraction of water	F_water	%	0.11	0.13	0.09	0.0008 - 0.66
Fraction of pasture	F_pasture	%	0.0011	0.066	0.10	0 - 0.59
Human impact index	H_index	n.a	0.148	0.237	0.248	0 – 0.97
Altitude	altitude	m	317	399.91	303.3	2 - 1555
30 days period solar radiation	solar_30days	J m^{-2} day^{-1}	535185	562097	6514	405257 - 721716
30 days period precipitation	precipitation_30days	m	1.05	1.07	0.54	0.04 - 2.93
Temperature	temperature	C degree	19.10	19.31	4.78	6.70 - 34.7
Humidity	humidity	%	80.90	78.63	19.5	0 - 100
Pressure	pressure	hPa	978.91	970.76	35.2	848.8 - 1032.1

n.a. no unit is applicable.

1.3.3 Laboratory analyses

Samples for the DOC and DIC, TN (mg L^{-1}) and TP (µg L^{-1}) were analyzed at the GRIL laboratory, located at the Université du Québec à Montréal. Surface samples for $\delta^{18}O_2$ were analyzed at the Ján Veizer Stable Isotope Laboratory at the University of Ottawa using standard methods (Barth et al. 2004). Samples for δ^2H and oxygen $\delta^{18}O_2$ in water were analyzed with an LGR (Los Gatos Research) model T-LWIA-45-EP Off-Axis Integrated Cavity Output Spectroscopy (OA-ICOS) at the GEOTOP laboratory, located at the Université du Québec à Montréal.

1.3.4 Calculation of lake metabolism

We estimated summertime mixed-layer metabolic rates in lakes using an oxygen isotope ($\delta^{18}O_2$) approach, which combines ambient O_2 concentration and its isotopic signature, as detailed in Bogard et al. (2017) and Bocaniov et al. (2012, 2015). This method was chosen because it offers a straightforward and efficient

way to provide an ecosystem-level estimate of metabolism, where samples could be collected within a single day. Other methods, which require extended incubation times or intensive sampling efforts, were not feasible for this large-scale study. This method accounts for the effects of air-lake gas exchange on the DO pool and due to its free-water nature, it does not distinguish between biotic and abiotic DO fluxes, therefore, the overall R rates presented here include several O_2 consumption pathways, such as pelagic R, organic matter photo-oxidation, and benthic metabolism (portion of sediments in contact with mixed layer). This method assumes steady state conditions, where there is no net change in either DO concentration or $\delta^{18}O_2$ over a diel cycle. As proposed by Bogard et al. (2017), we estimated volumetric GPP and R with the following equations:

$$GPP = \left(\frac{kO2}{Zmix}\right) x \frac{[(DO\ x\ (b-c)) - (DOsat\ x\ (a-c))]}{d-c} \quad (1)$$

$$R = \left(\frac{kO2}{Zmix}\right) x \frac{[(DO\ x\ (b-d)) - (DOsat\ x\ (a-d))]}{d-c} \quad (2)$$

where Z_{mix} is the mix layer depth, DO and DO_{sat} are the ambient and saturation DO (mg L^{-1}) concentrations, respectively, and k_{O2} (m d^{-1}) is the gas transfer coefficient of oxygen We calculated the atomic fractions for dissolved oxygen (AF_{DO}), water (AF_{H2O}), and atmosphere (AF_{atm}) following Bogard et al. (2020), based on Hotchkiss and Hall (2014) as in Equation 3, where *i* is the atomic fraction:

$$AFi = \frac{Rsample}{1+Rsample} \quad (3)$$

Where R_{sample} is the ratio of heavy to light samples ($^{18:16}O_{DO}$, $^{18:16}O_{atm}$, $^{18:16}O_{H2O}$, respectively). The parameters *a*, *b*, *c* and *d* in the equations (1) and (2) were calculated as follows: $a = AF_{atm} \times a_s \times a_g$, $b = AF_{DO} \times a_g$, $c = AF_{DO} \times a_c$ and $d = AF_{H2O} \times a_p$. We considered fractionation factors for gas exchange at the air-water interface (a_g = 0.9972, Knox et al. 1992), gas solubility in water (a_s = 1.0007, Benson & Krause, 1980), and DO production through photosynthesis (a_p = 1.000, Guy et al. 1993) and consumption of DO at ecosystem-level (a_c = 0.985, Bogard et al. 2017). We also estimated NEP as the difference between GPP and R, and calculated the ratios of GPP:R.

Out of the 742 lakes from which isotopic data were collected and DO saturation either collected (n = 672) or predicted (n = 70), 162 implausible negative values of GPP and of R (a common feature of free-water

and isotope methods) had to be excluded. The exclusion was based solely on sign coherence, removing values where GPP or R were negative and thus inconsistent with expected metabolic processes. Although this reduced the dataset size, we found no systematic pattern in the excluded data, suggesting that this exclusion was unlikely to introduce bias. These implausible values could be the result of compounded errors and uncertainty in sampling, preservation, laboratory analysis for isotope measurements and oxygen concentration measurements, uncertainty in gas exchange estimates, and/or physical mixing effects on gas solubility (Bocaniov et al. 2015), none of which can be accounted for a posteriori. While we assume that the retained values are robust, we acknowledge that they are all associated with some degree of uncertainty. Unfortunately, due to methodological constraints, we are unable to fully quantify this uncertainty.

1.3.5 Calculation of light penetration and mixed layer depth

To estimate water column light penetration, daily means of surface Photosynthetic Active Radiation (PAR) for the 15 days before sampling were obtained from CERES_SYN1deg_Ed4A (Wielicki et al. 1996) measurements. The ASTM G173-03 surface reference spectrum (a reference irradiance spectrum of direct Sunlight on Earth's surface at the ground level) (ASTM 2003) was used as the direct irradiance spectrum (incident irradiance only from the direction of the Sun), while the diffuse spectrum (incident irradiance from angles different than the direction of the direct sunlight, scattered by the atmosphere and clouds) was adjusted following Bartlett et al. (1998) with a cloud cover factor of 1. Day lengths and solar zenith angles (the angle between the local vertical (zenith) and the line connecting the observer to the Sun) were obtained from the R package Suncalc. The mean PAR was corrected for daytime only before averaging. Backscattering coefficient (bb) was measured for 73 lakes using HydroScat-6 (HOBI Labs, United States). The remainder was estimated using a boosted regression tree model that incorporated physicochemical and watershed variables, including *Chla*, TSS, and Secchi depth. The attenuation coefficient (k_d) was calculated following (Lee et al. 2005) and validated with in situ measurements obtained with the Hyperpro (Satlantic, Canada) and C-OPS (Biospherical Instruments, United States) profiling probes. Average epilimnion irradiance (E_{wc}) was calculated as the result of the integral calculation:

$$\widehat{E_{wc}} = \frac{\int_{Z_{mix}}^{z=0} E(z)\, dz}{Z_{mix}} \qquad (4)$$

which, after separately adding the direct and diffuse irradiances results in:

$$\widehat{E_{wc}} = \frac{E_{Dir}(1-e^{-k_d^{Dir}Z_{mix}})}{k_d^{Dir}Z_{mix}} + \frac{E_{Dif}(1-e^{-k_d^{Dif}Z_{mix}})}{k_d^{Dif}Z_{mix}} \quad (5)$$

Where E_{Dir} and E_{Dif} represent the direct and diffuse surface irradiance, k_d^{Dir} and k_d^{Dif} denote the direct and diffuse k_d values (calculated differently due to distinct incidence angles), and Z_{mix} is the mixing depth. We estimated the mixed layer depth (Z_{mix}) from temperature profiles and water density gradients (threshold 0.1 kg m^{-3} per meter) in cases of thermal stratification, we used the "Lake Parameters" add-in in JPM software (version Pro 16.0.0).

1.3.6 Numerical analysis

Due to probe malfunction, there were a number of sites (n = 70) for which ambient O_2 concentration values were missing, and we therefore could not estimate metabolism using the O_2 isotope balance. In order to be able to incorporate these sites into our metabolism database, we carried out a gap-filling procedure to estimate the ambient O_2 concentrations. Missing oxygen saturation data were predicted using the MissForest() function. MissForest is a random forest imputation algorithm for predicting missing data, as detailed in (Stekhoven and Bühlmann 2012). We assessed the accuracy of the method by randomly setting aside a portion of the observed data. To estimate the uncertainty associated with this method, we calculated the root-mean-square error (RMSE), mean absolute percentage error (MAPE), and Mean absolute error (MAE) of the predicted values (see Supplemental Figure S1-6). Deuterium excess (d-excess) values were calculated as follows: d-excess = $\delta^2H - 8 \times \delta^{18}O_2$ (Turner et al. 2014). d-excess represents a departure from the local meteoric signal and as such can be used as an index of the degree of evaporation in a given water sample.

Prior to estimating correlations and regressions, we conducted checks on the normality of metabolic and other environmental data. When necessary, we transformed the data to log10; for NEP, a log10 + X transformation was applied, and for fractional watershed variables, a square root approach was used. Statistical tests were conducted using R v. 4.1.0 (R Core Team 2021) and JMP (JMP.Pro 16.0.0). To explore the patterns of GPP, R, and NEP across Canada within the context of five drainage basins, we used boxplots, and we assessed statistical differences among the major watersheds using one-way ANOVA and Tukey's HSD post hoc tests. To investigate spatial patterns in environmental variables, we categorized the variables into three main groups based on land use, climate, and in-lake characteristics, and we then conducted a principal components analysis (PCA) on these three categories separately with the *ggfortify* package's

precomp() function. Additionally, we used the cluster package's *kmean*() function to predict cluster membership, which could shed light on the regional-scale context leading to geographic differences. We used silhouette coefficients to determine the appropriate number of clusters. To understand how metabolic components covaried with environmental variables, we employed redundancy analysis (RDA) using the vegan package's *rda*() function. To assess the relationship between metabolic rates and in-lake controls (e.g., TP) and the indicators of metabolism (e.g., *Chla*), we binned the data, which helped to extract large-scale patterns and smooth out the noise in these patterns, and we then performed regression analysis on each relationship. Based on the RDA and regressions on binned data, we developed models for GPP, R, and NEP to assess regional differences in the relationship between metabolic rates and environmental drivers. We performed an analysis of covariance (ANCOVA) using a linear regression model using the *lm*() function from the stats package. In each ANCOVA model, our model included GPP, R and NEP as the dependent variable, region (continental basin) as a categorical predictor, and environmental variables that were most strongly related to each of the metabolic components as covariates (e.g., TP, TN). We initially fit models with interaction terms, allowing slopes to vary by region. If the interaction term was not significant ($p < 0.01$), we refit the model without interaction effects, assuming a common slope across regions while allowing intercepts to vary. In addition, we used the *lmer*() function from the *lme4* package to fit linear mixed-effects models and the *ranef*() function to extract random effects. Standard errors of the model coefficients were obtained using the *se.coef*() function from the arm package. In each model, the fixed effects were environmental variable (e.g., TP, TN), while the random effect was region (continental basin). We compared predictive power of fits among linear mixed-effect based on AIC, LogLik and deviance in Table S1.2, versus ANCOVA models that included same response and predictor variable. We extracted intercepts and slopes from better models to identify the strongest regional drivers.

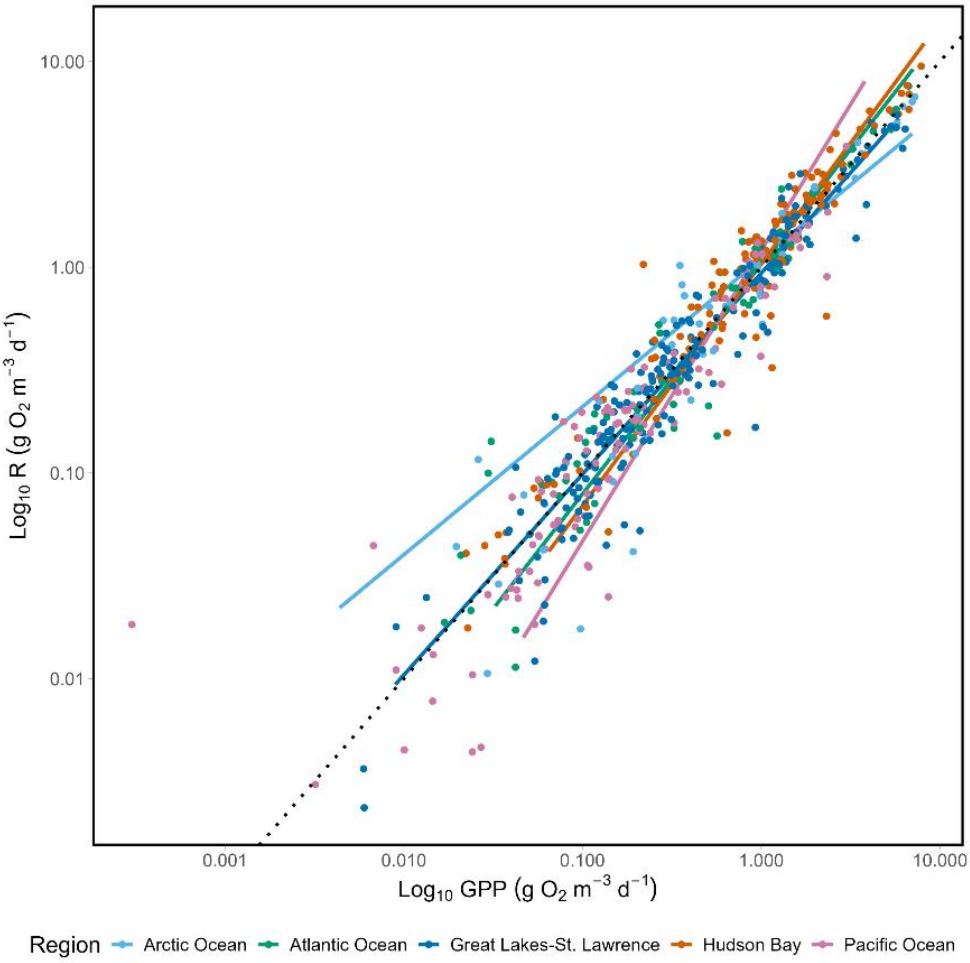

Figure 1.2: Relation between $\log_{10}$ R (g O_2 m^{-3} d^{-1}) and $\log_{10}$ GPP (g O_2 m^{-3} d^{-1}) for lakes across all sampled lakes. The dashed line represents the 1:1 line, while the solid lines represent the regression lines.

1.4 Results

1.4.1 Metabolic rates vary greatly across Canadian lakes

Across all lakes, summer GPP and R ranged 4 orders of magnitude, from 0.003 to 6.8 g O_2 m^{-3} day^{-1}, and 0.003 to 7.5 g O_2 m^{-3} day^{-1}, for GPP and R, respectively, and there was a strong positive relationship between both (log slope of 0.99; Figure 1.2). The log slopes for the subset of lakes for the different basins mostly followed the overall slope, but diverged for the Arctic Ocean basin (log slope of 0.98) and for the Pacific Ocean basin (log slope of 0.89) (Figure 1.2). For a small subset of the lakes in Eastern Ontario (n = 8) we had seasonal data gathered over 2 consecutive years, and the temporal patterns of GPP and R in

these lakes are presented in Figure S1.1, S1.2 in the Supporting Information section. Rates of GPP remained well constrained around an overall mean of 0.77 g O_2 m^{-3} day^{-1} (range between 0.008 and 1 g O_2 m^{-3} day^{-1}) in this subset of mostly oligotrophic boreal lakes from spring to fall and between years. There were a handful of isolated spikes in GPP that were several fold above in magnitude to this range of values, and which mostly occurred in the same 2 lakes, and are largely unexplained. Likewise, R varied within a relatively narrow range between lakes, seasons and years (mean 1.22 g O_2 m^{-3} day^{-1}, range between 0.02 and 1 g O_2 m^{-3} day^{-1}), and it too had several unexplained isolated spikes in two of the lakes. Overall, these results suggest that water column metabolism is constrained within fairly narrow ranges in these boreal lakes, at least within the open water season, and that our approach effectively captures this range of variability.

Figure 1.3 shows the magnitude and variability of summer lake GPP, R, NEP, and GPP:R within and across the five major Canadian continental basins. The highest median volumetric GPP was observed in the Hudson Bay basin (Prairies and Boreal Plain regions) and the lowest was observed in the Pacific Ocean basin (Montane Cordillera and Pacific Maritimes regions), (1.10, 0.14 g O_2 m^{-3} day^{-1}, respectively), and the same pattern was observed for R (1.15, 0.15 g O_2 m^{-3} day^{-1}). NEP spanned a much narrower range than GPP and R (4-fold) from -2 to 1.9 g O_2 m^{-3} day^{-1}. The overall ANOVA for GPP, R, NEP, and GPP:R was statistically significant, but post hoc analysis revealed that not all regions were significantly different from each other for GPP (ANOVA: $p < 0.0001$, post hoc: Pacific Ocean > Hudson Bay, Great Lakes-St. Lawrence, Arctic Ocean, Atlantic Ocean; Hudson Bay > Great Lakes-St. Lawrence; Figure 1.3a), R (ANOVA: $p < 0.0001$, post hoc: Pacific Ocean > Hudson Bay, Great Lakes-St. Lawrence, Arctic Ocean, Atlantic Ocean; Hudson Bay > Great Lakes-St. Lawrence; Figure 1.3b), NEP (ANOVA: $p < 0.0001$; post hoc: Pacific Ocean > Hudson Bay; Hudson Bay > Great Lakes-St. Lawrence, Arctic Ocean; Figure 3c), and GPP:R (ANOVA: $p < 0.01$, post hoc: Pacific Ocean > Hudson Bay; Figure 1.3d).

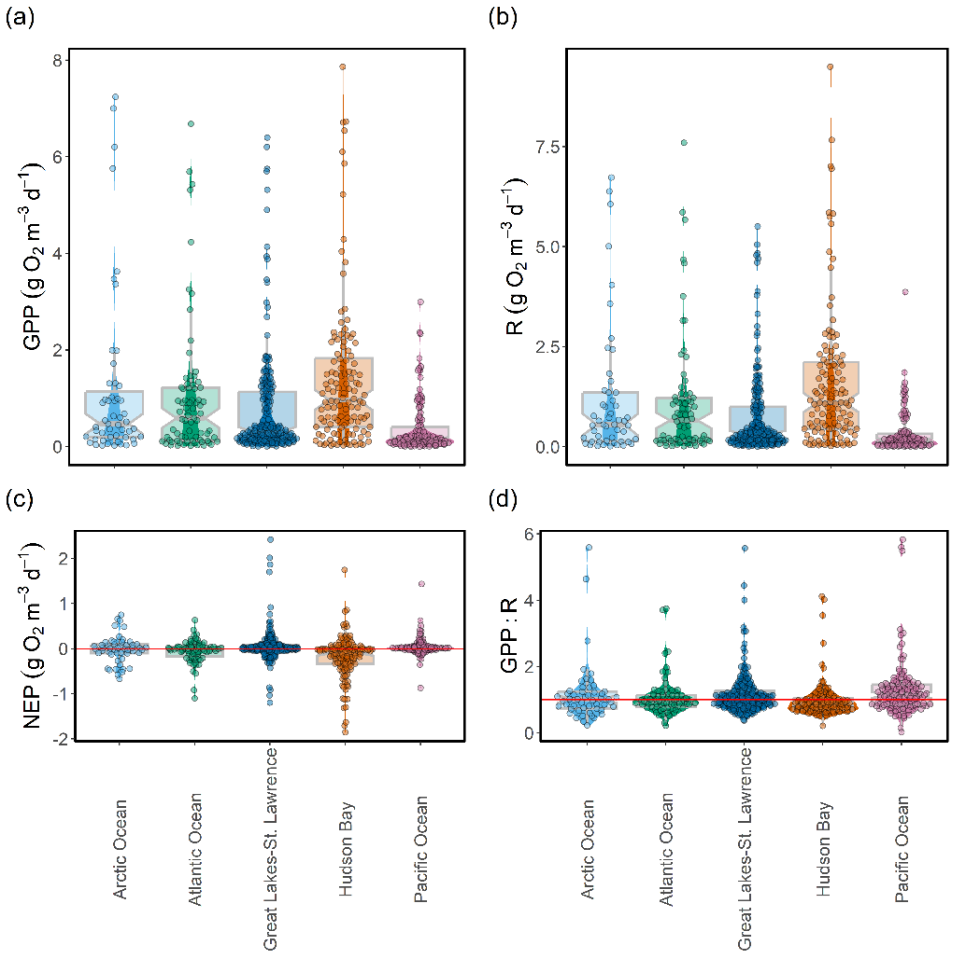

Figure 1.3: Volumetric rates of gross primary production (GPP; a), respiration (R; b), net ecosystem production (NEP; c), and the ratio between GPP and R (GPP:R; d), Red line in panels (c) and (d) indicates the threshold between autotrophic and heterotrophic status

1.4.2 Large gradients in land use, climate and in-lake properties across regions

The studied lakes and their watersheds covered a wide range of land use, climate, and lake properties, with large differences both between and within the five major Canadian continental basins, as shown in Table 1.1 Lakes were located within watersheds ranging in size from 0.2 to 37,459 km^2. The lakes also ranged greatly in size and morphometry, ranging from 1 to 64 m in maximum depth (Z_{max}), and from 0.002 to 4345 km^2 in surface area, and had a wide range of physicochemical and trophic conditions, including DOC concentrations varying from 1 to 220 mg L^{-1}, chlorophyll concentrations ranging from 0.04 to 72 µg L^{-1}, TP from 2 to 99.8 µg L^{-1}, TN from 0.01 to 11.9 mg L^{-1}, and Secchi depth varying from 0.07 to 15.4 m. These diverse characteristics influenced the PCA results (Figure. 1.4), where the PCA that corresponds to in-lake

features (Figure. 1.4a) distinguished four clusters of lakes, with nutrients, DOC and d-excess weighing strongly on PCA1 separating clusters 1 and 2, and catchment and lake area, water temperature and light weighing on PCA2 and further separating clusters 3 and 4. Figure 1.4b shows the PCA distribution of lakes based on climate and geographic variables, which resulted in two clusters of lakes, with latitude, altitude and longitude weighing strongly on PCA1 and separating clusters 1 and 2, and precipitation and humidity and solar radiation weighing on PCA2. Figure 1.4c shows the PCA distribution of lakes based on land use and human impact variables, which resulted in four clusters of lakes, with fraction of agriculture, pasture and natural lands weighing strongly on PCA1, separating clusters 2 and 3, and fraction of urban and water weighing on PCA2 and separating clusters 1 and 4. Cluster 3 in the PCA (a) corresponds geographically to cluster 3 in the PCA (c), where there is a high human impact and a large fraction of agriculture and pasture. In these regions lakes typically have high concentrations of DOC, TP and TN.

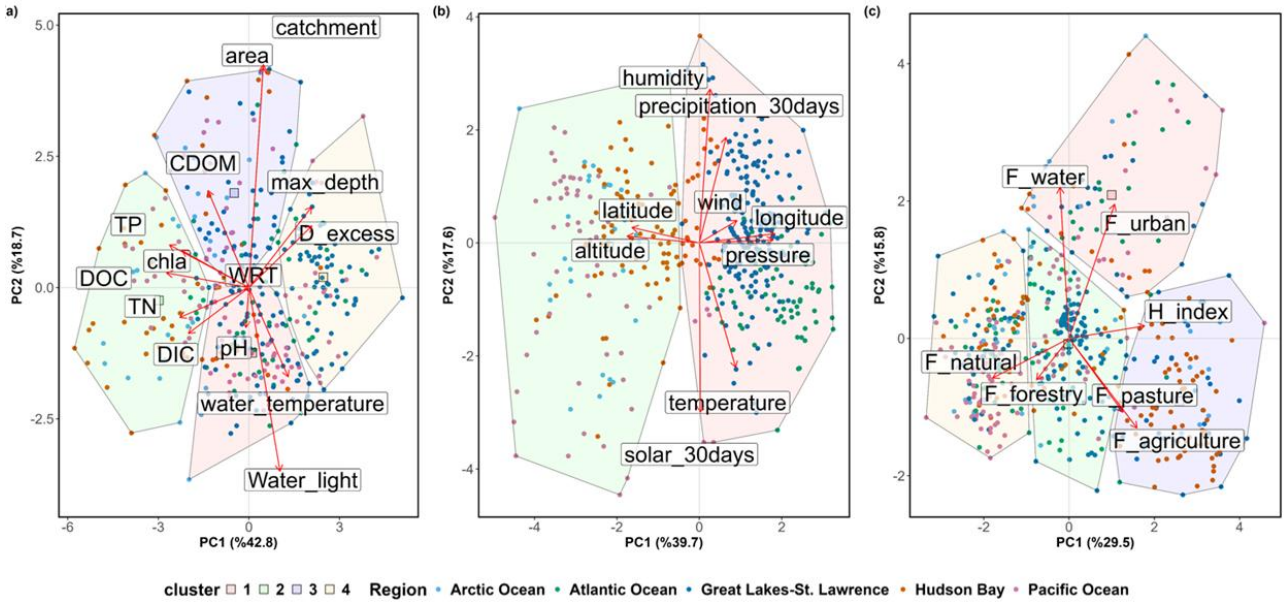

Figure 1.4: Principal component analysis (PCA) on three different groups of environmental variables (a) in-lake (b) climate, (c) land use and human impact for the sampled lakes (dots), clusters were added and shown as polygons. Abbreviations explained in Table 1.1.

1.4.3 Relationship between metabolism and chemistry, climate and land use properties

In the RDA that corresponds to in-lake characteristics, both GPP and R plotted on the negative side of RDA1, closely aligning with potential drivers such as TP, TN, DOC as well as the indicator variable *Chla* (Figure. 1.5). NEP and GPP:R plotted on the second axis of the RDA positioned near additional controls such as light and temperature (Figure. 1.5). Figure 1.5b shows the RDA distribution of lakes based on

climate and geographic variables, and here GPP and R plotted on axis RDA2 close to latitude and altitude, whereas NEP and GPP:R weighed on RDA1 and close to solar radiation and atmospheric pressure (Figure. 1.5b). Figure 1.5c shows the RDA based on land use and human impact variables, and here GPP and R plotted on RDA1 and close to controls including fraction of agriculture, fraction of pasture, and human impact index, whereas NEP and GPP:R plotted on RDA2 close to fraction of forest. Results from Figure 1.5a and Figure 1.5c shows that rates of metabolism relate strongly with nutrients, which are linked to land use and land cover. The results of regression analysis on binned data show a strong positive relationship between lake GPP and R, and *Chla*, TP, TN, and DOC (Figure S1.4, S1.5).

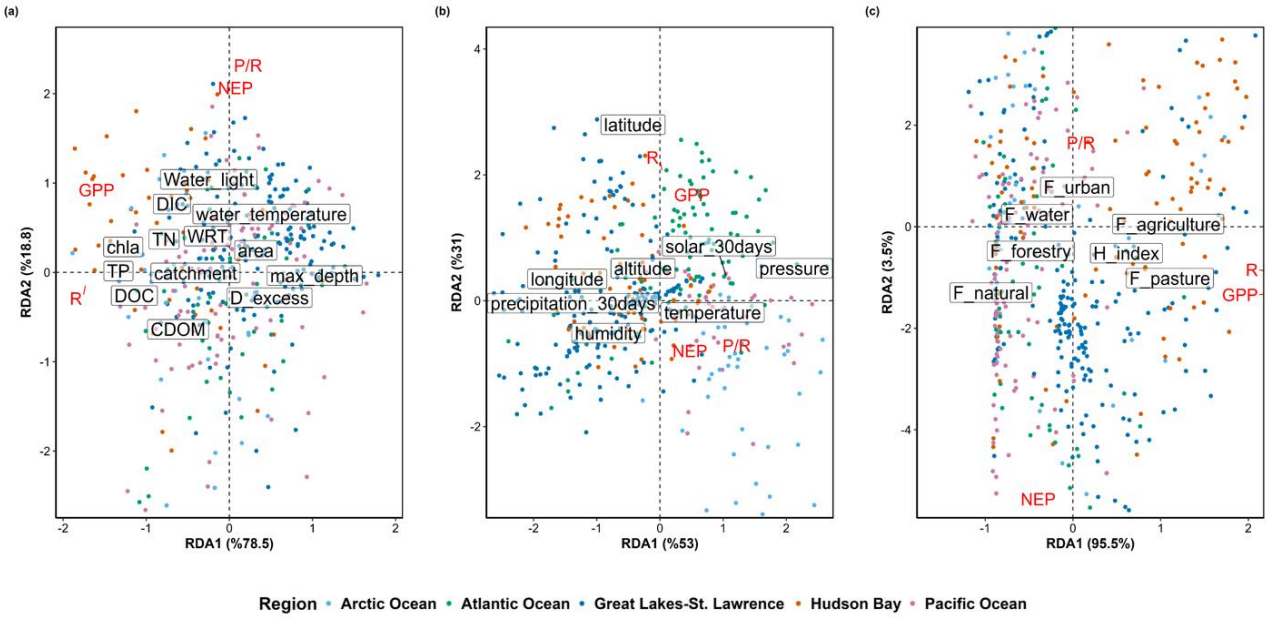

Figure 1.5: Ordination plot of redundancy analysis (RDA) of GPP, R, NEP, and GPP:R with three different groups of environmental variables (a) in-lake (b) climate, (c) land use and human impact. Abbreviations explained in the Table 1.1.

We developed ANCOVA models for GPP, R, and NEP that incorporated the variables that were most strongly related to each of the metabolic components. For GPP and R, the resulting significant models were structured as functions of TP, TN, DOC as well as *Chla* ($p < 0.001$) (Figure. 1.6 and 1.7, Table 1.2 and S1.1). In contrast, the models for NEP were a function of water column light and CDOM ($p < 0.001$) (Figure 1.8 and Table 1.2, S1.1). In the case of GPP and R, introducing a random slope did not enhance the explained variation and did not significantly differ from the models with variable intercepts only. The only exception was the model for GPP as a function of TN, where the interaction between the region and TN was significant ($p < 0.01$). Consequently, we retained a model with varying slopes in this specific case. The

models, as presented in Figures. 1.6 and 1.7 and Table 1.2 and S1.1, reveal considerable variation in the intercepts (in addition to the slopes concerning GPP-TN), of the relationships between nutrients, DOC, *Chla*, and both GPP and R across the different regions. For two models of NEP, region had a significant interaction with both CDOM and light ($p < 0.001$), and we therefore retained the slopes (Figure 1.6, Table 1.2 and S1.1). Across all models, the marginal R^2, which represents the variance explained by the fixed effect alone, was consistently lower than the conditional R^2 (Table S1.3). This pattern indicates that incorporating regional clustering as a random effect improved model performance by capturing additional variation in the response variable.

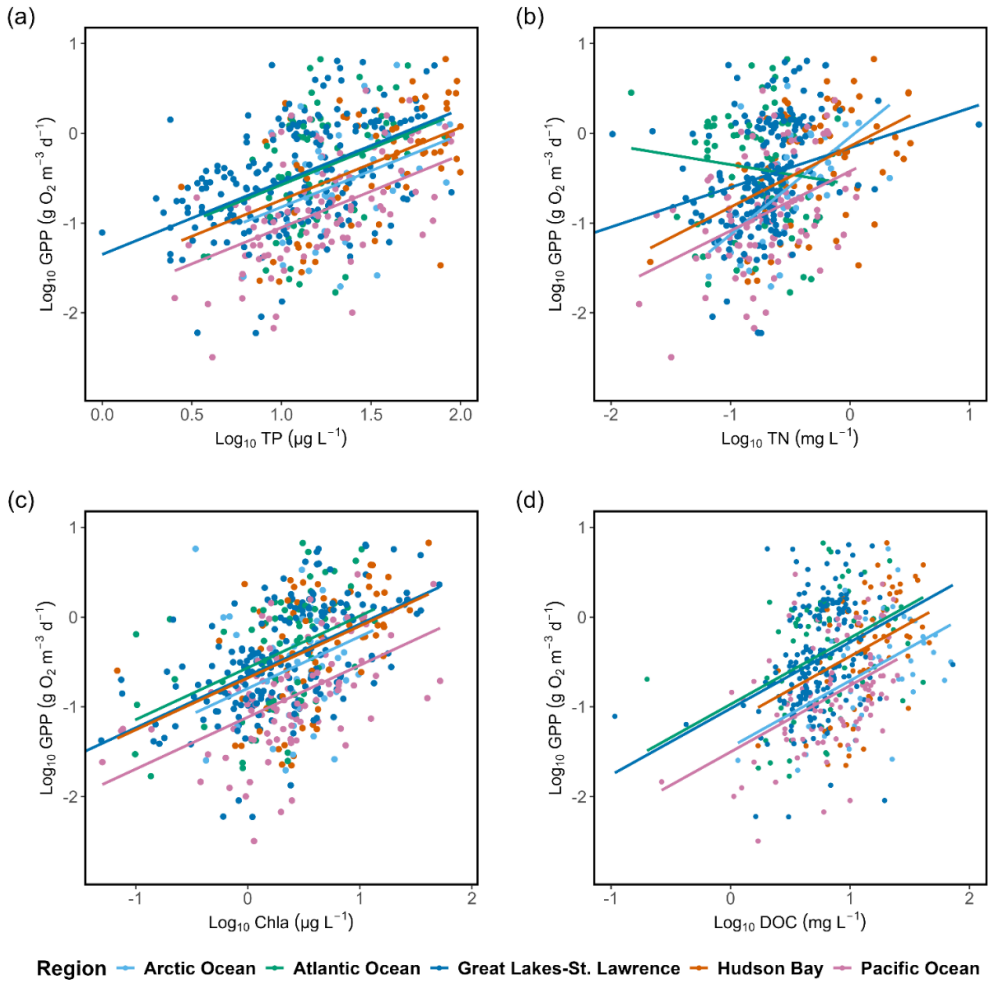

Figure 1.6: Relationships between GPP and DOC, TP, TN, and *Chla* across regions.

1.4.4 Local and regional drivers interact to shape metabolic rates

We further assessed if the regional intercepts and slopes in the case of GPP vs TN as well as of the NEP vs CDOM and light models described in the previous section were related to regional properties, including average regional topography, land-use and hydrology. The intercepts of the GPP and R models were negatively related with the average regional altitude, as shown in Figure 1.9. Additionally, the region-specific slopes of the GPP vs TN and NEP vs CDOM and also NEP vs light models were significantly negatively related with the average regional d-excess, as shown in Figure 1.9.

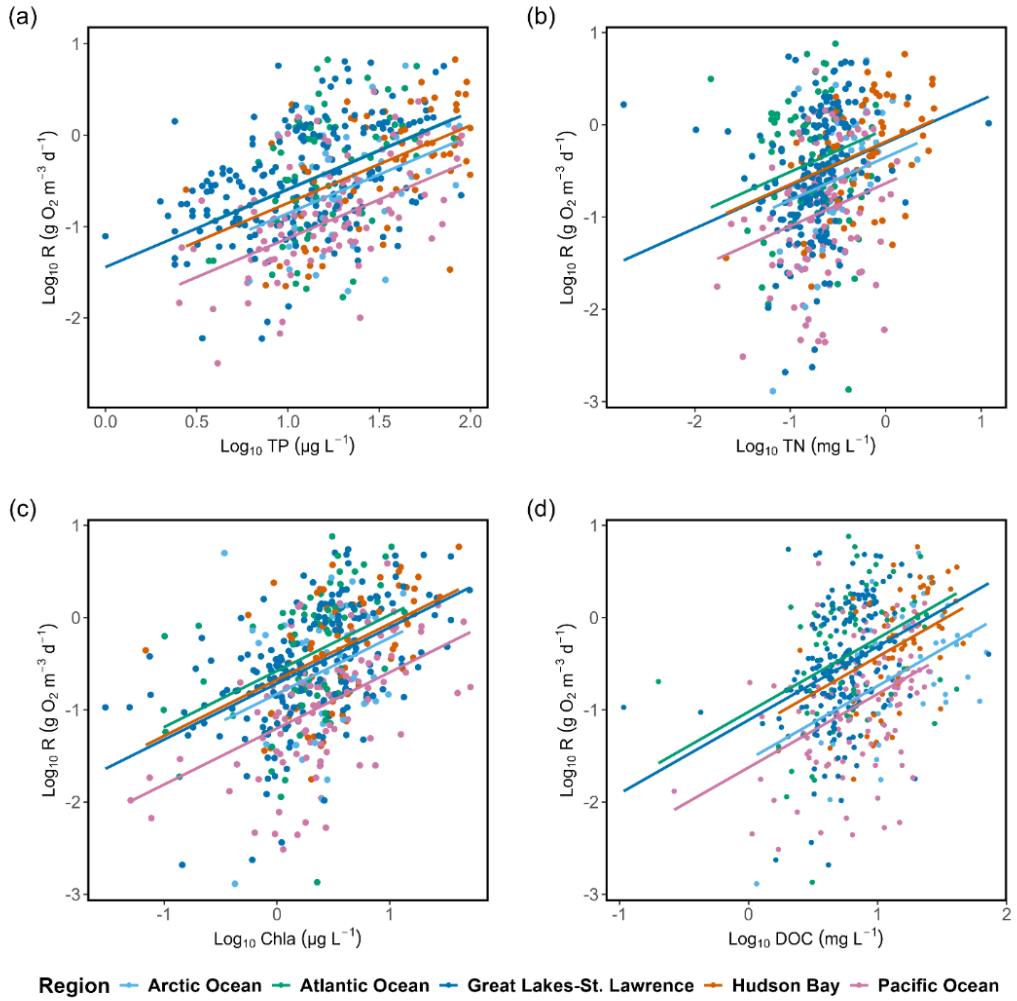

Figure 1.7: Relationships between R with TP, *Chla*, TN and DOC across different regions.

1.5 Discussion

Our study identified the most critical variables explaining summer lake metabolism at a continental scale, and we have further that demonstrated that the effect of these potential drivers on GPP and R is modulated by regional topographic and hydrologic properties. Importantly, it would not have been possible to detect these regional effects if we had only studied lake metabolism at a local scale. We have shown that the proportion of lakes that are net heterotrophic during the summer across the continent is around %50, and they also suggest that the drivers of lake NEP at macro scales may be very different from those of GPP and R. GPP and R vary by four orders of magnitude in lakes across the Canadian landscape, but the ensuing net metabolic balance (NEP) was much more narrowly constrained, ranging only by four fold, suggesting a very tight coupling between primary production and respiration across these systems. The primary drivers of GPP and R during summer were DOC and secondarily nutrients (TP and TN), whereas lake morphometry, light and temperature did not appear to have a major influence at this scale. NEP, on the other hand, had very different drivers, and was positively related to light and negatively related to CDOM. These patterns of lake metabolism as a function of environmental drivers were modulated by average regional properties, such as mean altitude and average lake water isotopic composition (d-excess), and this generates regional differences in both the regional baseline lake metabolism and the response of lakes to core drivers. The continental scale patterns in lake metabolism therefore emerge from the interaction between local drivers that shape the metabolism of individual lakes, and landscape level drivers that integrate regional properties and that modulate the collective response of lakes to the environment and climate.

1.5.1 Lake metabolism varies strongly across regions

Despite a growing number of metabolism studies, cross-regional explorations of lake metabolism are still scarce. Most previous studies have either been carried out within a single watershed or region (Bogard et al. 2020; Smits et al. 2024) or global but with extremely low spatial resolution (Hoellein et al. 2013; Holgerson et al. 2022), and in this regard, our study bridges these scales because it focuses at the continental scale with a relatively high degree of spatial resolution. This approach enables us to capture regional variability and contextual differences that are often overlooked in more localized or small site number or coarse-scale analyses. The results obtained from the ANCOVA models have revealed significant variation in the region-specific intercepts of the models of GPP and R as a function of TP, chlorophyll and DOC, and also the slopes of the GPP vs TN model. In addition, there were region-specific differences in the slopes of models of NEP vs CDOM and light. These relationships suggest that for any given level of TP, DOC,

or *Chla*, lakes in different regions had different average GPP and R, and different responses to changes in TN. Likewise, lakes in different regions had different responses in terms of NEP to light and CDOM, and therefore, these variables alone cannot account for the differences in average lake GPP, R and NEP across the five major continental basins in Canada. These variations in the intercepts (and slopes) of these regional models suggest the influence of other environmental variables at the regional scale that interact with local drivers to modulate cross-regional patterns in lake metabolism.

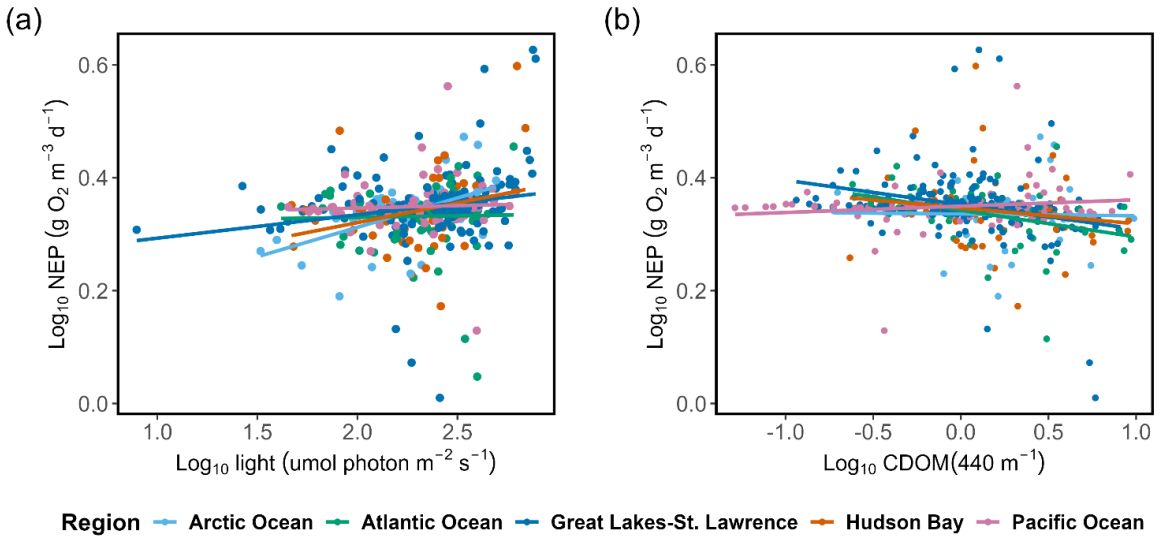

Figure 1.8: Relationships between NEP with CDOM and water column light across lakes in different lakes.

In this regard, we observed a significant negative relationship between the extracted regional intercepts of the GPP vs TP and DOC models and average regional altitude, suggesting that the baseline metabolism of lakes varies systematically as a function of the regional topography. Previous studies carried out mostly at the watershed scale have also emphasized the importance of landscape or network location in determining key aspects of lake function, and converged to suggest that metabolism (both GPP and R) should tend to increase at the lower portions of the basins and at lower elevations, due to increased terrestrial influence (Ask et al. 2012; Ogdahl et al. 2010; Seekell et al. 2018; Puts et al. 2023). In our study, and in agreement to Vanderwall et al. (2024) at the individual lake scale, altitude did not have a significant effect on metabolism, but it did influence the average differences in lake metabolism that we observed between regions, suggesting that it likely acts as an integrative proxy for other topographic, morphologic, land use and climatic variables. Interestingly, temperature did not appear to drive either the local patterns

in metabolism, nor the cross regional differences in these patterns, contrary to what has been shown in other studies (Karlsson et al. 2005), but this may reflect a rather narrow temperature range that was encountered during the summer sampling window despite the large latitudinal and longitudinal spatial coverage of the study. It is very likely, however, that temperature effects may also be integrated within altitude. In addition, altitude may influence the partitioning of total ecosystem GPP into pelagic and benthic components, as has been shown in several studies (Puts et al. 2023; Ayala-Borda et al. 2024), and this could further influence our estimates of ecosystem GPP and R, since our approach may not effectively capture a portion of benthic metabolism. In the case of the GPP vs TN model, not only the intercept, but also the slope was related to d-excess, and this would suggest that lakes in regions that are characterized by higher mean d-excess are more sensitive to changes in nutrient loading than those in lower altitude regions in terms of primary production.

The cross-lake patterns in NEP also were modulated regionally. NEP was negatively related to CDOM and positively related to water column light, and the region-specific slopes of these relationships were positively related to the average regional water d-excess, a proxy for the degree of evaporation of lake water relative to precipitation. Lakes within the Hudson Bay basin, for example, had on average lower d-excess values, implying lake waters that have been more influenced by the effects of evaporation, and therefore likely subjected to longer overall residence times and processing of materials (Ayala-Borda et al., 2024). The negative relationship between the slope of the NEP vs CDOM model and the average regional d-excess would suggest that lakes in regions with high average d-excess (and likely low average water residence time in the network), such as the Great Lakes-St. Lawrence and Atlantic Ocean basins, are more sensitive to changes in CDOM, and therefore to browning, in terms of their net metabolism. This agrees with recent studies that have shown that long term trends in DOC, CDOM and lake CO_2 concentrations are most tightly coupled in lakes with short residence time within the boreal landscape (Rodríguez-Cardona et al. 2023).

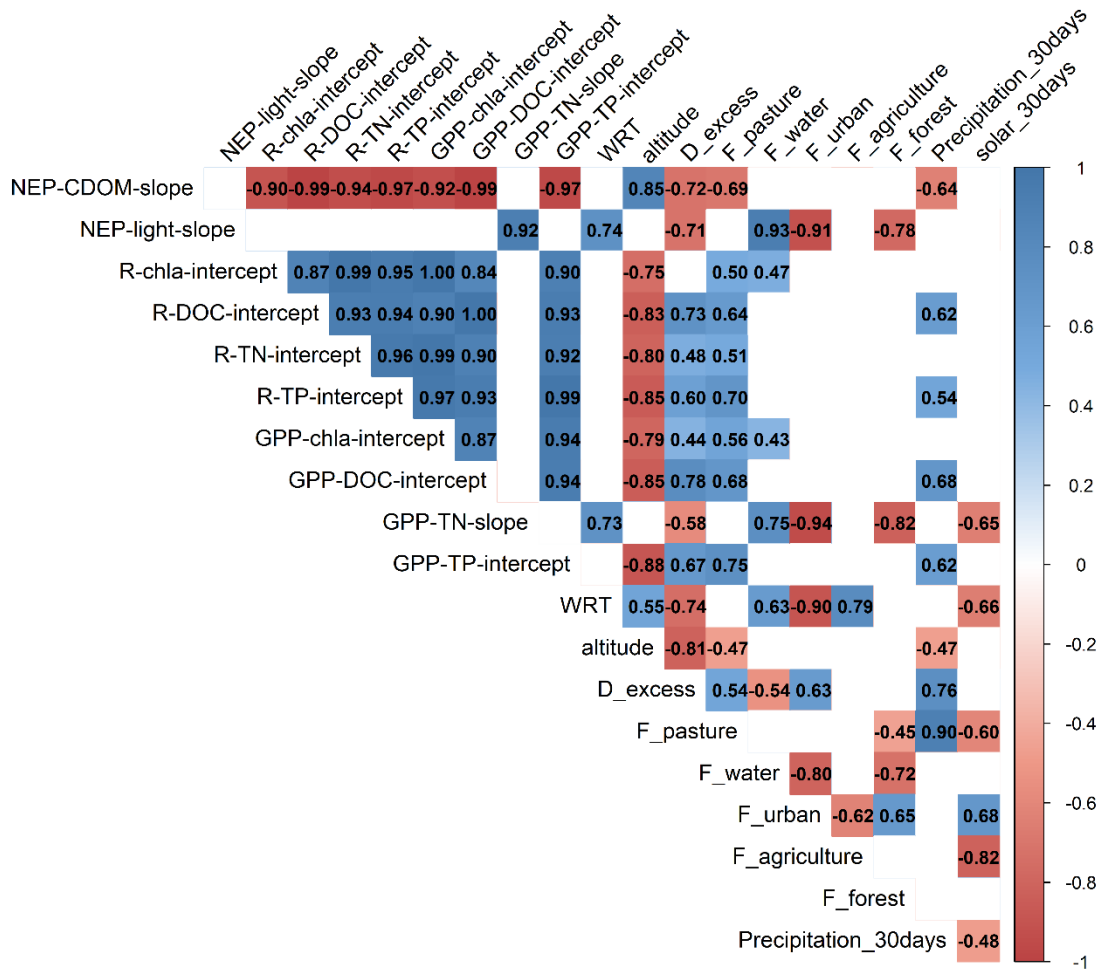

Figure 1.9: Correlation matrix intercepts/slope from models predicting GPP, R and NEP and the relationship with average regional properties.

1.5.2 Relationship between GPP and R at macroscale

The GPP in our study lakes ranged from 0.003 to 6.8 g O^2 m^{-3} day^{-1}, whereas R ranged from 0.003 to 7.5 g O^2 m^{-3} day^{-1}, within the reported range for lakes in temperate and boreal regions (Solomon et al. 2013; LaBuhn and Klump 2016; Bogard et al. 2020; Barbosa et al. 2023; Corman et al. 2023; Lu et al. 2023) from studies that used oxygen isotopes or high-frequency free gas concentrations methods to estimate metabolism. There was a strong relationship between GPP and R across Canadian lakes (Figure. 1.2), consistent with previous studies (Solomon et al. 2013; Oleksy et al. 2021; Klaus et al. 2022; Corman et al. 2023; Ayala-Borda et al. 2024). This relationship further suggests that lakes across Canada are roughly evenly distributed on both sides of the one-to-one line along the entire productivity range. These findings somewhat contradict with earlier studies, such as del Giorgio et al. (1994), which had concluded that

oligotrophic lakes tended to be net heterotrophic, and that lakes tend to be increasingly net autotrophic as they become more productive. Our results, however, suggest that at a continental scale, whether a lake is heterotrophic or autotrophic is not solely a function of its trophic status or productivity, and that other drivers must influence the metabolic balance, at least during summer.

This strong relationship between GPP and R (Figure. 1.2) during the summer window suggests a close coupling between primary production and respiration, and that most of what is produced by pelagic (or benthic in the case of shallow lakes) autotrophs is readily mineralized. As a result, there is a relatively small range of variation in the net balance between GPP and R (NEP) across lakes and regions, with most values falling between -0.5 to 0.5 g O^2 m^{-3} $day^{-1}$and an overall positive average across all lakes of 0.003 g O^2 m^{-3} day^{-1}, despite the over four orders of magnitude range in both GPP and R across these lakes; this pattern of relatively narrow NEP that is evenly spread around neutrality across lakes that differ greatly in absolute metabolic rates has been observed before (Klaus et al. 2022; Corman et al. 2023), and has consequences for various aspects of lake and landscape functioning, including CO_2 dynamics and emissions, as well as organic carbon burial in sediments. In theory, lakes with positive NEP should be net CO_2 sinks, at least during the periods of net autotrophy, and conversely, they should act as net CO_2 sources when NEP is < 0. Previous studies, however, have shown that surface water CO_2 dynamics are not always tightly coupled to metabolism (Vachon et al. 2020; Bogard et al. 2016), and that lakes that are net autotrophic can still be supersaturated in CO_2 and emit the gas to the atmosphere (Bogard and del Giorgio 2016) due among other to groundwater inputs. Likewise, the fact that over half of the lakes sampled appeared to have negative NEP (GPP < R) does not imply that these lakes do not bury organic matter in the sediments, because it has been shown that a significant fraction of the sediment accumulation in northern lakes is driven by the sedimentation of terrestrially derived organic matter (Gudasz et al. 2017), and this pathway likely occurs across all lakes and is decoupled from both GPP and R.

Our results show that whereas roughly half of all lakes (50.1%) appear to be net heterotrophic during the summer, there are strong regional differences in the distribution of the net metabolic balance. Lakes in the Great Lakes-St. Lawrence (55%), Pacific Ocean (63%), and Arctic Ocean (51%) tended to be net autotrophic (NEP = GPP > R), whereas the majority of lakes in the Atlantic Ocean (55%) and Hudson Bay (71%) watersheds were predominantly net heterotrophic (NEP = GPP < R). Correspondingly, the GPP:R ratio was predominantly <1 in the Atlantic Ocean and Hudson Bay, whereas GPP:R was on median >1 in the Arctic Ocean, Great Lakes-St. Lawrence, and Pacific Ocean. Other studies have shown that about two-

thirds of boreal lakes sampled were net heterotrophic (Bogard & del Giorgio 2016), with a similar proportion for northern US lakes (Mcdonald et al., 2013). On the other hand, Ayala-Borda et al. (2024) showed that about 90% of the lakes in a sub-arctic region were net autotrophic, and Bogard et al. (2019) showed summertime autotrophy dominated lakes of interior Alaska. Clearly, net metabolism (at least during summertime) appears to have important regional structure, albeit with considerable variability of NEP within a given region.

Table 1.2: Analysis of variance results for different models used in this study.

Model	Variable	F value	Pr(>F)
GPP-TP	Log10 TP	106.78	<0.001***
	Region	13.33	0.001***
GPP-TN	Log10 TN	43.78	0.001***
	Region	9.71	0.001***
	TN*Region	4.91	0.001***
GPP-DOC	Log10 DOC	64.62	0.001***
	Region	13.96	0.001***
GPP-*Chla*	Log10 *Chla*	126.45	<0.001***
	Region	13.27	0.001***
R-TP	Log10 TP	121.30	<0.001***
	Region	16.07	0.001***
R-TN	Log10 TN	39.69	0.001***
	Region	12.03	<0.001***
R-DOC	Log10 DOC	76.78	0.001***
	DOC	16.86	<0.001***
R-*Chla*	Log10 *Chla*	135.57	0.001***
	Region	16.96	0.001***
NEP-CDOM	Log10 CDOM	13.71	0.001***
	Region	0.98	0.4147
	CDOM*Region	4.42	0.0016**
NEP-light	Log10 light	20.23	0.001***
	Region	1.57	0.17974
	Light*Region	2.19	0.06852.

Significant : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1

1.5.3 Main drivers of ecosystem metabolism

Our results show a strong positive relationship between lake GPP and R, and DOC as well as proxies of trophic status such as TP, TN (Suppl. Fig S1.4, S1.5). Previous studies have shown that GPP and R in lakes vary as a function of nutrient status (i.e., TP, TN) (Holgerson et al. 2022; Corman et al. 2023), and our study confirmed that TP had a stronger influence on both GPP and R (prior to binning: R^2 = 0.20, R^2 = 0.19, respectively; after binning R^2 = 0.66, R^2 = 0.69) compared to TN (prior to binning: R^2 = 0.13, R^2 = 0.10, respectively; after binning R^2 = 0.70, R^2 = 0.65). Other studies have concluded that TN had a stronger effect

on GPP in some northern landscapes (Bogard et al. 2020), suggesting that these landscapes may be a mosaic of patches of N and P limitation. As expected, GPP and R in our study showed a strong positive correlation with its indicator, *Chla* (prior to binning: R^2 = 0.20, R^2 = 0.19, respectively; after binning R^2 = 0.94, R^2 = 0.90, Suppl. Fig S1.4 and S1.5), in agreement with previous studies (Forget et al. 2009; Solomon et al. 2013). DOC, on the other hand, may influence GPP in various ways, negatively by altering the light climate and competing with algae for light, positively by delivering nutrients and increasing temperature (Puts et al. 2023), potentially establishing complex, unimodal relationships along gradients of concentration (Puts et al. 2023; Seekell et al. 2015).

The strong coupling between GPP and R suggests that GPP-derived organic matter serves as the main substrate for R across lakes (Figure. 1.2), yet DOC still plays a major role in modulating lake metabolism. We found DOC was positively related to both GPP (prior to binning: R^2 = 0.11; after binning: R^2 = 0.75) and R (prior to binning: R^2 = 0.13; after binning: R^2 = 0.65) along the entire range of DOC concentrations, with no evidence of unimodal relationships, or of a declining GPP at high DOC concentrations, in agreement with previous studies (Bogard et al. 2020; Ask et al. 2012). This would suggest that the enhancement due to the release from nutrient limitation associated to DOC delivery may overwhelm the light inhibition effect on GPP (Solomon et al. 2015; Kelly et al. 2018; Rivera Vasconcelos et al. 2018). Indeed, this is reflected in our RDA, where TP, TN and DOC are aligned with each other and strongly with GPP. Although R responds positively to the same drivers as GPP, the strength of these relationships is not the same for GPP and for R, this decouples GPP from R, and this is reflected in patterns in NEP. NEP was positively correlated with water column light, likely due to its enhancement of GPP, and was negatively correlated with CDOM, which may reflect both reduced light availability (Solomon et al., 2013; Puts et al., 2023) and a potential increase in respiration through the provision of substrates (del Giorgio and Peters, 1994). This in turn highlights the central yet complex role that DOC (and the associated CDOM) similar to (Oleksy et al. 2024) plays in these systems: The strong positive relationships between DOC and both GPP and R indicate that allochthonous organic matter inputs significantly influence lake metabolism across lakes, by delivering nutrients but also by acting as a substrate for R. Yet the enhancement of R by DOC and the depression of GPP by CDOM appear to exceed the positive effect of DOC on GPP, leading to reduced NEP. As a result, the most negative NEP values we observed were in highly colored, DOC-rich lakes, likely reflecting this overall balance of effects allochthonous organic matter on lake metabolism.

Lakes across the Northern Hemisphere are undergoing major changes at multiple levels, in terms of nutrient loading, browning, hydrology, and warming and shifting physical regimes, among others (Woolway et al. 2022). Here we have shown that summer lake metabolism is modulated by a complex interplay between these factors, such that the outcome of future changes in terms of net ecosystem metabolism is still difficult to predict. In addition, we have shown that there is a regional aspect to these changes, with lakes in some regions seemingly more sensitive to eutrophication or to browning. There are significant gaps that still need to be filled in our understanding of lake metabolism at the continental scales, however. A better representation of northern boreal and subarctic regions is critically needed, and studies should incorporate the extremes of the lake size distribution, since the largest lakes and the smallest water bodies play major roles in the landscape yet are systematically under sampled. In addition, future studies should extend these summer results to a full annual cycle, in order to capture other key periods such as spring. Improving our sampling and measurement strategies will contribute to our capacity to predict the direction and magnitude of future change in lake metabolism.

1.6 Acknowledgements

We thank the lead PI of NSERC-funded LakePulse Project, Yannick Huot, as well as the many technicians, students, and research professionals involved in the collection and generation of data for the network. We would like to thank Bruno Cremella for providing data and calculations for light in the water column. We would like to thank the Interuniversity Research Group in Limnology (Groupe de recherche interuniversitaire en limnologie [GRIL]) and their funders, the Fonds de recherche—nature et technologie (FRQNT, Québec). This is part of the research program of the NSERC / HydroQuébec Research Chair in Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBas) led by PdG and which also supported AS. MB was supported by the Canada Research Chairs Program and the University of Lethbridge. We declare no conflicts of interest.

1.7 Data availability

All the raw data that supports the findings of this paper will be made entirely available to the public as part of the LakePulse database (https://lakepulse.ca/national-lake-pulsedatabase/).

1.8 Supplementary Information

1.8.1 Representativeness of summer sampling

Our extensive study delved into the dynamics of ecosystem metabolism across a diverse spectrum of lakes throughout Canada, encompassing a wide array of environmental variables. We used the $\delta^{18}O_2$ isotopic free-water approach to explore how drivers interact over spatial scales to shape lake metabolism across the continent. We acknowledge, however, the limited temporal coverage of our study, but emphasize the difficulty of carrying macroscale comparative studies together with detailed temporal assessments for a large number of lakes across many regions, in terms of logistics, costs and resources. The limitation of summer sampling also has been discussed in detail in Bogard et al. (2020). There is nevertheless a certain degree of temporal variability embedded in the database and in the relationships that we show, since samples were collected during a 2-month window from July to September, over three consecutive years, but this variability is mostly at a regional scale, not at the individual lake scale. The limited seasonal data that we gathered (Supplementary Figure S1.1, S1.2) would suggest that at least in a subset of oligotrophic boreal lakes, the patterns of GPP and R remained relatively stable from spring to fall and over two consecutive years, similar to patterns demonstrated in Bogard et al. (2020) and Ladwig et al. (2022), suggesting that our conclusions may extend beyond the summer window, but this may not be the case for other types of lakes, for example, eutrophic lakes in the Prairies, which may have much more pronounced seasonal metabolic patterns (Ladwig et al. 2022).

1.8.2 Figures and tables

Table S1.1: Results of different ANCOVA models for GPP, R and NEP.

Model	Region	Intercept	Slope	P-value
GPP-TP	Arctic Ocean	-1.58	0.79	0.001***
	Atlantic Ocean	-1.35	0.79	0.39*
	Hudson Bay	-1.54	0.79	0.74
	Great Lakes St. Lawrence	-1.34	0.79	0.015*
	Pacific Ocean	-1.79	0.79	0.034*
GPP-TN	Arctic Ocean	-0.049	1.03	0.78
	Atlantic Ocean	-0.59	-0.24	0.001***
	Hudson Bay	-0.14	0.68	0.517
	Great Lake St. Lawrence	-0.17	0.43	0.023*
	Pacific Ocean	-0.37	0.67	0.062.
GPP-DOC	Arctic Ocean	-1.44	0.79	0.001***
	Atlantic Ocean	-0.97	0.79	0.001***
	Hudson Bay	-1.17	0.79	0.016*
	Great Lakes St. Lawrence	-1.01	0.79	0.001***
	Pacific Ocean	-1.50	0.79	0.618
GPP-*Chla*	Arctic Ocean	-0.79	0.57	0.001***
	Atlantic Ocean	-0.56	0.57	0.027*
	Hudson Bay	-0.67	0.57	0.27
	Great Lakes St. Lawrence	-0.65	0.57	0.12
	Pacific Ocean	-1.11	0.57	0.0012**
R-TP	Arctic Ocean	-1.70	0.84	0.001***
	Atlantic Ocean	-1.43	0.84	0.019*
	Hudson Bay	-1.59	0.84	0.332
	Great Lakes St. Lawrence	-1.44	0.84	0.01*
	Pacific Ocean	-1.97	0.84	0.01*
R-TN	Arctic Ocean	-0.35	0.46	0.001***
	Atlantic Ocean	-0.045	0.46	0.0144*
	Hudson Bay	-0.182	0.46	0.178
	Great Lakes St. Lawrence	-0.195	0.46	0.152
	Pacific Ocean	-0.634	0.46	0.0161
R-DOC	Arctic Ocean	-1.53	0.79	0.001***
	Atlantic Ocean	-1.02	0.79	0.001***
	Hudson Bay	-1.22	0.79	0.0082**
	Great Lakes St. Lawrence	-1.11	0.79	0.001***
	Pacific Ocean	-1.62	0.79	0.451
R-*Chla*	Arctic Ocean	-0.82	0.60	0.001***
	Atlantic Ocean	-0.57	0.60	0.026*
	Hudson Bay	-0.67	0.60	0.181
	Great Lakes St. Lawrence	-0.71	0.60	0.249
	Pacific Ocean	-1.20	0.60	0.001***
NEP-CDOM	Arctic Ocean	0.333	-0.00013	0.001***
	Atlantic Ocean	0.342	-0.0472	0.95
	Hudson Bay	0.340	-0.0021	0.54
	Great Lakes St. Lawrence	0.352	-0.0433	0.065.
	Pacific Ocean	0.349	0.011	0.14
NEP-light	Arctic Ocean	0.098	0.106	0.15
	Atlantic Ocean	0.275	0.025	0.0708.
	Hudson Bay	0.180	0.069	0.373
	Great Lakes St. Lawrence	0.251	0.0415	0.0444*
	Pacific Ocean	0.329	0.0085	0.0076**

Significant : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table S1.2: Results of comparing ANCOVA models vs Mixed effect models for GPP, R and NEP.

Model	AIC	LogLik	Deviance	Chisq	Pr(>F)
GPP-TP					
Mixed effect model	806.16	-399.08	798.16		
ANCOVA model	795.75	-390.88	781.75	16.40	<0.001***
GPP-TN					
Mixed effect model	866.69	-427.35	854.69		
ANCOVA model	851.89	-414.94	829.89	24.80	<0.001***
GPP-DOC					
Mixed effect model	840.85	-416.43	832.85		
ANCOVA	829.31	-407.65	815.31	17.55	<0.001***
GPP-*Chla*					
Mixed effect model	789.32	-390.66	781.32		
ANCOVA	778.37	-382.19	764.37	16.95	<0.001***
R-TP					
Mixed effect model	870.95	-431.48	862.95		
ANCOVA	860.33	-423.17	846.33	16.62	<0.001***
R-TN					
Mixed effect model	940.89	-464.44	928.89		
ANCOVA	934.75	-460.38	920.75	8.13	0.004**
R-DOC					
Mixed effect model	900.43	-446.21	892.43		
ANCOVA	888.59	-437.29	874.59	17.83	<0.001***
R-*Chla*					
Mixed effect model	855.19	-423.59	847.19		
ANCOVA	843.73	-414.86	829.73	17.45	<0.001***
NEP-light					
Mixed effect model	-1365.0	688.48	-1377.0		
ANCOVA	-1369.6	695.81	-1391.6	14.66	0.011*
NEP-CDOM					
Mixed effect model	-1380.6	696.31	-1392.6		
ANCOVA	-1384.4	703.22	-1406.4	13.81	0.016*

Significant: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table S1.3: Fixed effects of the linear mixed-effects models used to predict GPP, R and NEP (±Standard Error).

Variable	Fixed effects	Fixed intercept	R^2 marginal	R^2 Conditional
Log 10 GPP	0.803 ± 0.072 × log10 (TP)	-1.547 ± 0.131	0.20	0.29
Log 10 GPP	0.534 ± 0.197 × log10 (TN)	-0.251 ± 0.087	0.10	0.22
Log 10 GPP	0.576 ± 0.048 × log10 (*Chla*)	-0.761 ± 0.099	0.21	0.31
Log 10 GPP	0.727 ± 0.081 × log10 (DOC)	-1.209 ± 0.132	0.16	0.27
Log 10 R	0.845 ± 0.076 × log10 (TP)	-1.626 ± 0.141	0.20	0.30
Log 10 R	0.540 ± 0.226 × log10 (TN)	-0.286 ± 0.118	0.08	0.23
Log 10 R	0.607 ± 0.051 × log10 (*Chla*)	-0.800 ± 0.111	0.20	0.31
Log 10 R	0.786 ± 0.086 × log10 (DOC)	-1.291 ± 0.143	0.15	0.28
Log 10 NEP	0.042 ± 0.016 × log10 (light)	0.244 ± 0.038	0.04	0.06
Log 10 NEP	-0.024 ± 0.012 × log10 (CDOM)	0.348 ± 0.002	0.03	0.06

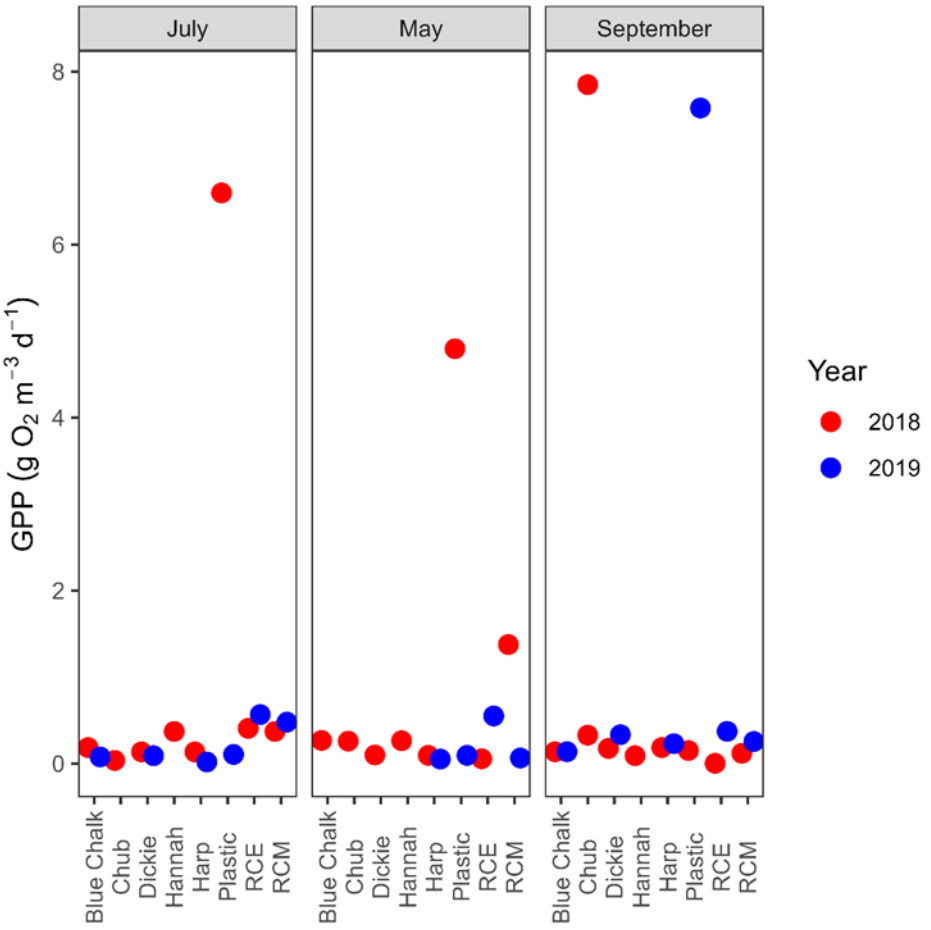

Figure S1.1: Boxplot showing temporal pattern of volumetric rates of GPP. In 2018 and 2019 in subset of lakes in May, July and September.

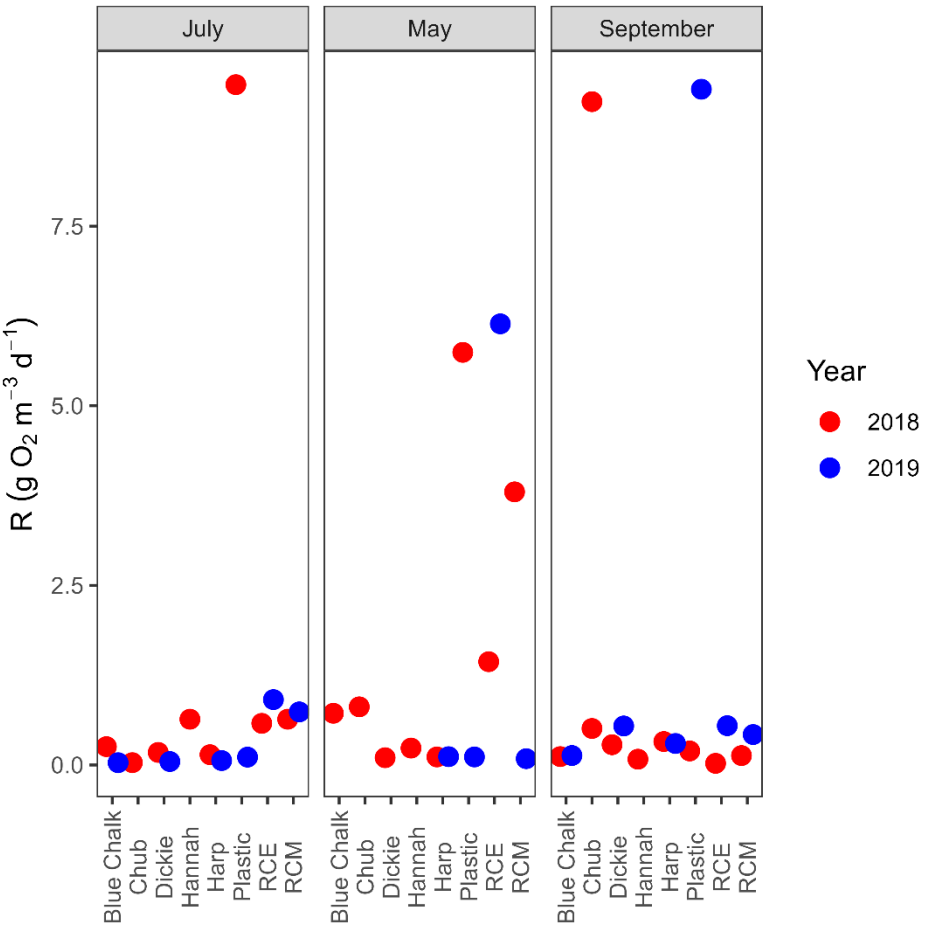

Figure S1.2: Boxplot showing temporal pattern of volumetric rates of R. In 2018 and 2019 in subset of lakes in May, July and September.

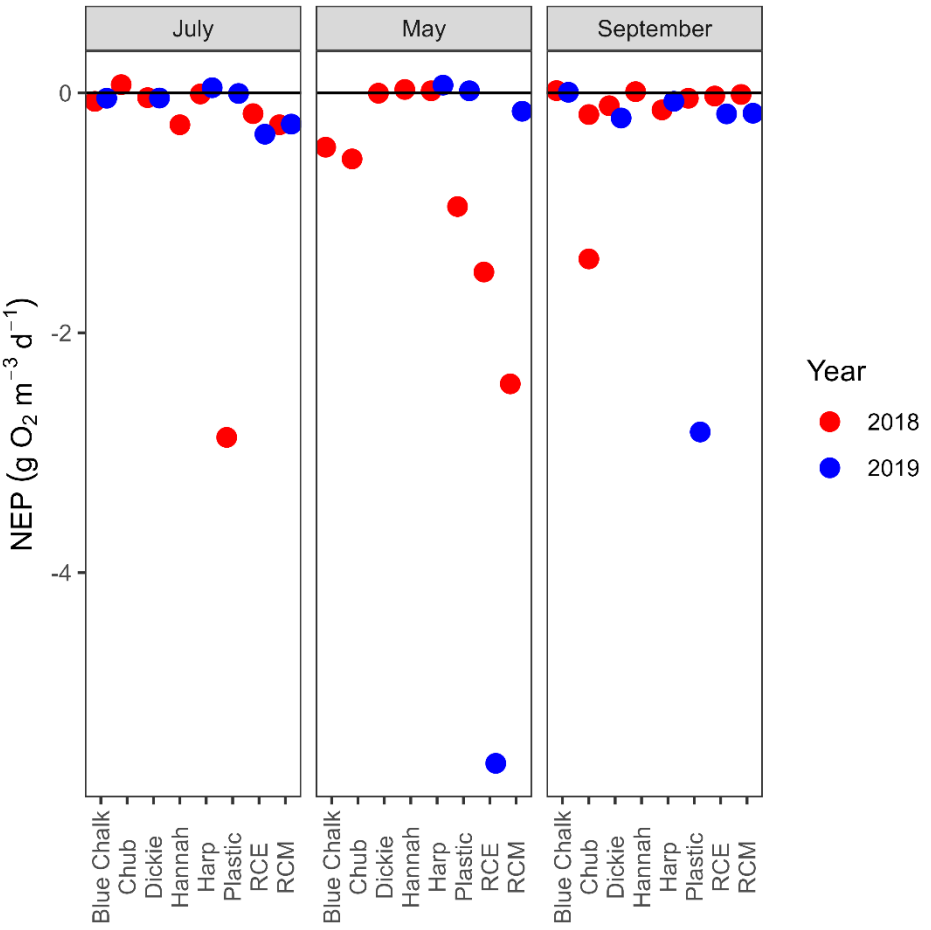

Figure S1.3: Boxplot showing temporal pattern of volumetric rates of NEP. In 2018 and 2019 in subset of lakes in May, July and September. Black line in panel indicates the threshold between autotrophic and heterotrophic status.

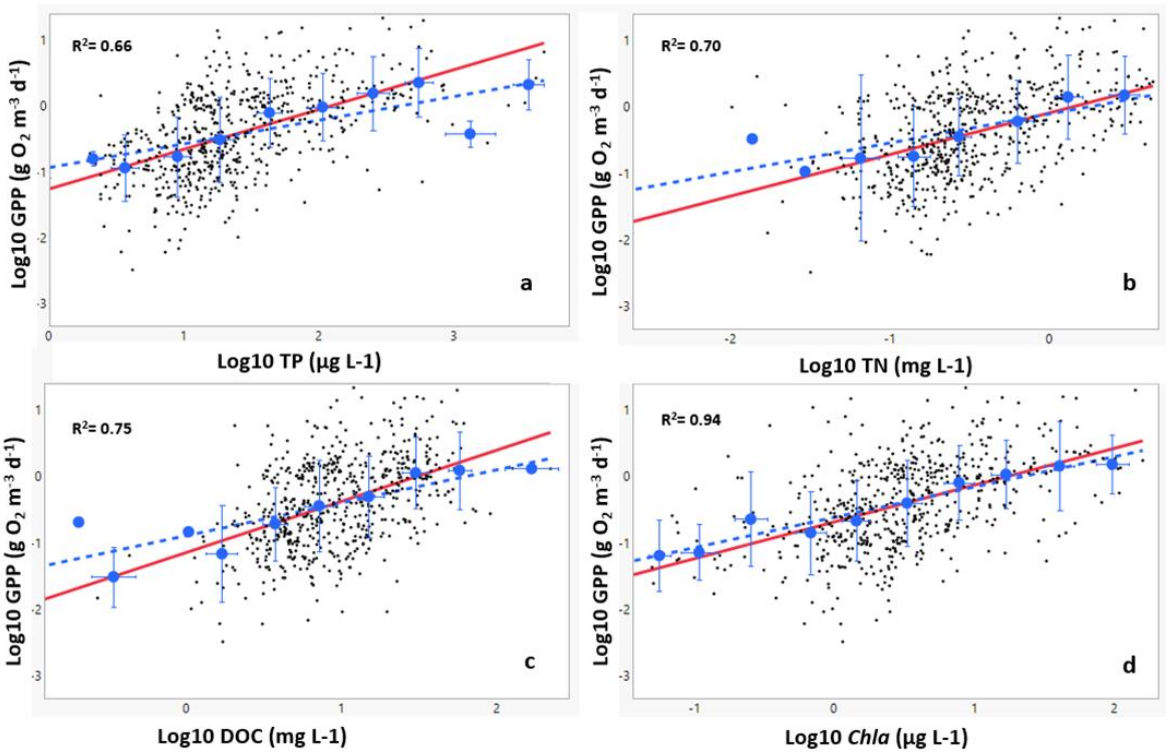

Figure S1.4: Relationships between binned log10 transformed GPP and TP, TN, DOC, *Chla*. The red line indicated fit line of original data. Blue line indicates fit line of binned data. Blue dots indicate Error bar on bins.

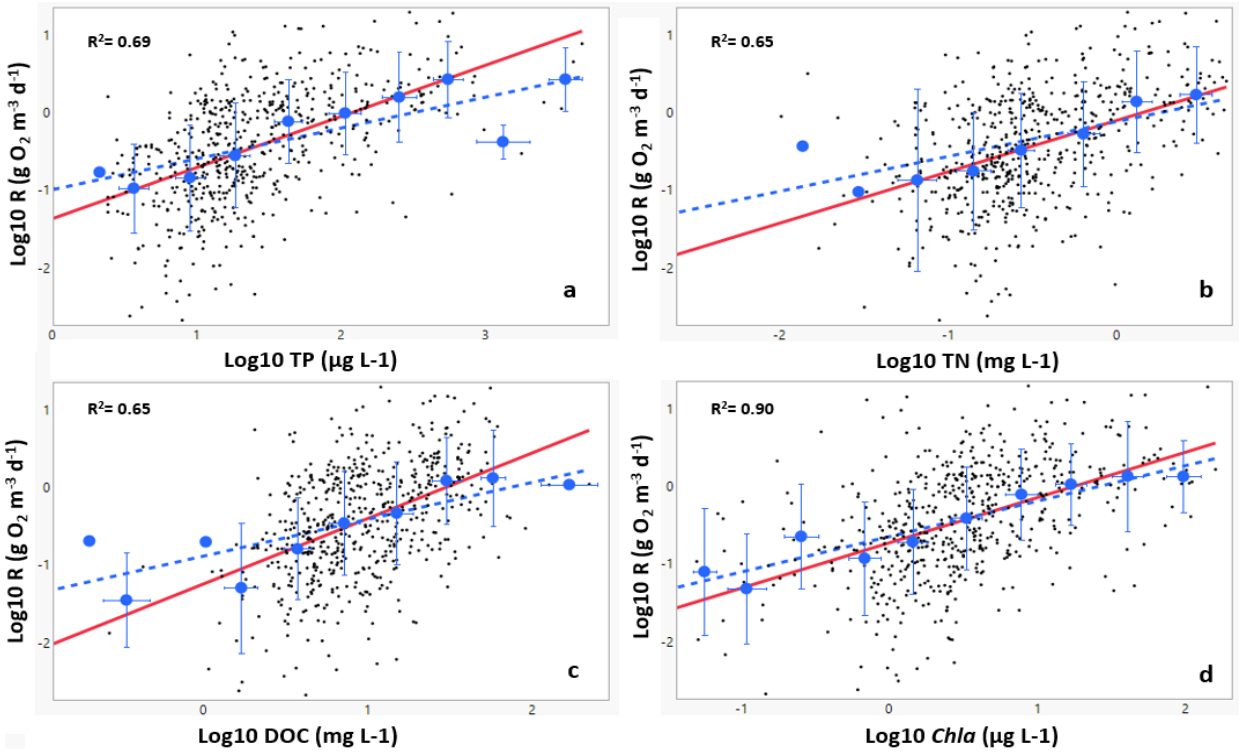

Figure S1.5: Relationships between binned log10 transformed R and TP, TN, DOC, *Chla*. The red line indicated fit line of original data. Blue line indicates fit line of binned data. Blue dots indicate Error bar on bins.

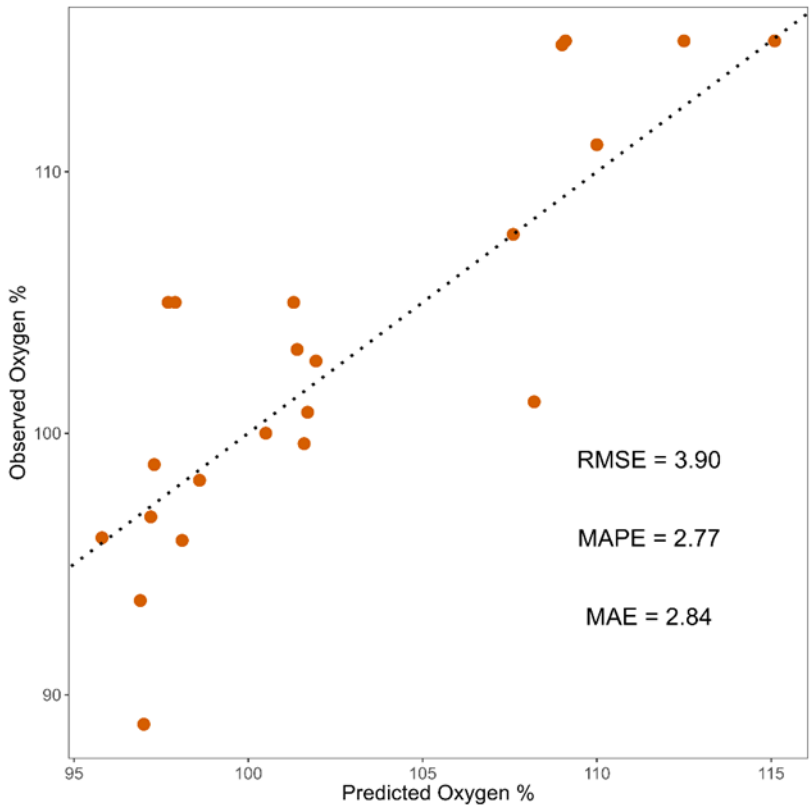

Figure S1.6: Relationships between observed and predicted oxygen%. Root-mean-square error (RMSE), mean absolute percentage error (MAPE), and Mean absolute error (MAE).

CHAPITRE 2

Hydrology and trophic status control lake dissolved organic matter concentration and composition at a continental scale

Amir Reza Shahabinia [1], Ryan. H. S. Hutchins[2], Jean-François Lapierre[3,4] and Paul A. del Giorgio[1]

[1] Department of Biological Sciences, University of Quebec at Montreal, Montreal, Quebec, Canada.

[2] Department of Chemistry and Biology, Toronto Metropolitan University.

[3] Department of Biological Sciences, University of Montreal, Montreal, Quebec, Canada.

[4] Groupe de Recherche Interuniversitaire en Limnologie

Keywords: Dissolved organic carbon (DOC), Dissolved organic matter (DOM), molecular composition, hydrology, continental scale, optical component

N.B. References cited in this are presented at the end of this thesis.

2.1 Abstract

Dissolved organic matter (DOM) is a key component of the lake biogeochemistry. Hydrology links variables influencing lake DOM at local and watershed scales, but its role at macroscales remains less understood. We studied the DOM concentration and composition from 548 lakes across the five major Canadian continental basins, using absorption spectroscopy and parallel factor analysis, and ultra‐high resolution mass spectroscopy, and linked this to deuterium excess (d-excess), derived from stable water isotopes as a proxy for evaporation, water residence time, and regional hydrology. DOM concentration and composition varied greatly within and across basins, with strong correlations between molecular and optical properties. At a continental scale, d-excess and TP concentration were the main drivers of DOM concentration and composition. TP positively influenced DOM concentration, and specific DOM components (e.g., Aliphatics), suggesting nutrient-driven effects on lake metabolism that varied regionally. DOM concentration declined with d-excess, but the relationships between individual DOM molecular composition classes and d-excess differed among components and basins, resulting in regional differences in DOM composition along hydrologic gradients. The inferred source composition DOM based on these patterns had subtle regional differences, with Aliphatics related to the average regional altitude, and Aromatics related to the average regional soil organic content. We show that DOM processing along the hydrologic continuum is the key factor establishing differences in DOM composition in lakes at a continental scale. Overall, TP influenced DOM through effects on primary production and metabolism, whereas d-excess, integrated the selective degradation and accumulation of DOM along the aquatic network.

2.2 Introduction

Dissolved organic matter (DOM) is a key limnological variable that influences multiple aspects of the physical, biogeochemical and ecological functioning of inland waters (Williamson et al., 1999; Prairie, 2008). Both the concentration and composition of DOM vary greatly among lakes and rivers within a given region, as a function of system morphometry, network positioning, trophic status, and local watershed features (Catalán et al., 2016; Weyhenmeyer & Conley, 2017) and also across regions, reflecting large scale patterns in climate, land-use and soil properties (Kothawala et al., 2014; Lapierre et al., 2015). The chemical composition of ambient DOM, which is an extremely complex mixture of tens of thousands of different molecules, is the net result of multiple pathways of production, transformation and degradations (Cooper et al., 2022) but it is also itself a driver of aquatic ecosystem processes (Tanentzap et al., 2019).

In this regard, the DOM observed in a given lake reflects the ambient production and transformation processes within the lake itself, and the loading of DOM that originates from the aquatic network feeding the lake, which itself has multiple origins and processing, including interactions with the surrounding landscape (Solomon et al., 2015). A significant portion of the DOM found in northern lakes originates from the surrounding landscapes, transported from soils to streams, rivers and eventually lakes via groundwater and surface runoff. During this transport, DOM undergoes extensive transformation and degradation, which alter its concentration as well as its optical and chemical properties (Ryan et al., 2024). The molecular composition of terrestrially derived DOM depends on landscape characteristics, such as vegetation type and soil organic content (Herreid et al., 2025), and is also shaped by regional hydrology, which regulates not only the movement of materials from land to water (Kurek et al., 2024), but also the transformation and persistence of DOM within aquatic networks (Liu et al., 2022). Once in the aquatic network, in-stream and in-lake processes further transform and degrade DOM, and generate new forms of DOM (Stadler et al., 2020; Orlova et al., 2024). Additionally, aquatic metabolism and primary production within lakes further influence DOM dynamics, and therefore factors such as trophic status and water residence time (WRT) have been shown to significantly influence DOM concentrations and its molecular and optical properties (Johnston et al., 2020; Pugh et al., 2021; Kurek et al., 2023).

Both in-lake and regional processes that influence lake DOC concentrations and DOM molecular composition are shaped by a multitude of environmental, geographic, and climatic variables (Ward et al., 2017). However, hydrology underlies most of these processes, acting as a common denominator that links the variables influencing lake DOM (Jones et al., 2024). The concentration and composition of DOM in any

given lake are influenced not only by the lake's WRT and recent regional precipitation (Kellerman et al., 2014; Xenopoulos et al., 2021) but also by the past hydrologic history of the aquatic network that is associated to the lake and its interactions with the surrounding landscape (Pugh et al., 2021). Whereas regional precipitation data can be easily obtained, and aspects of aquatic network hydrology, such as lake WRT and local runoff can be indirectly derived by combining remote sensing and modeling, it is much more difficult to derive an integrative perspective of the history of the water within the watershed, that includes the origin, pathways, and climatic and evaporative processes the water has experienced before and during its residence within the lake, all of which are highly relevant to DOM dynamics. In this regard, stable water isotopes integrate climatic signatures with hydrologic processes (Dee et al., 2023), offering insights into key factors influencing lake DOM dynamics at a regional and continental scale, including terrestrial DOM export and in-lake processing. Heavy and light isotopes undergo phase transitions at different rates due to variations in atomic mass, diffusivity, and vapor pressure (Dansgaard, 1964) and their ratio provides an integrated record of moisture exchange (Galewsky et al., 2016). This isotopic signature, with deuterium excess compared to oxygen δ18O2, captures the movement of water between land, sea, and atmosphere, shaping the fundamental states of the biosphere, ocean, and climate (Ines Fung, 2008). This integrative perspective from isotopic signatures is particularly valuable for understanding both the amount and the molecular composition of DOM that is found in lakes of different types and in different regions.

To better understand regional patterns in lake DOM, it is essential to place these within the broader context of large-scale hydrologic patterns. Further, understanding these links between DOM and hydrology is increasingly pressing as regional hydrologic dynamics are particularly vulnerable to climate and environmental change. Here we present a continental scale study of lake DOM, where we assessed the patterns of concentration and optical and molecular composition of DOM of lakes across Canada. The optical properties of DOM were evaluated using parallel factor analysis (PARAFAC), and the molecular properties of DOM were evaluated using Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS), in a wide range of lake types along broad climatic, land-use and hydrologic gradients. In particular, we assessed how large-scale patterns in hydrology, as revealed by water isotopes, modulate the influence of local and regional variables on continental scale patterns in lake DOM amount and composition.

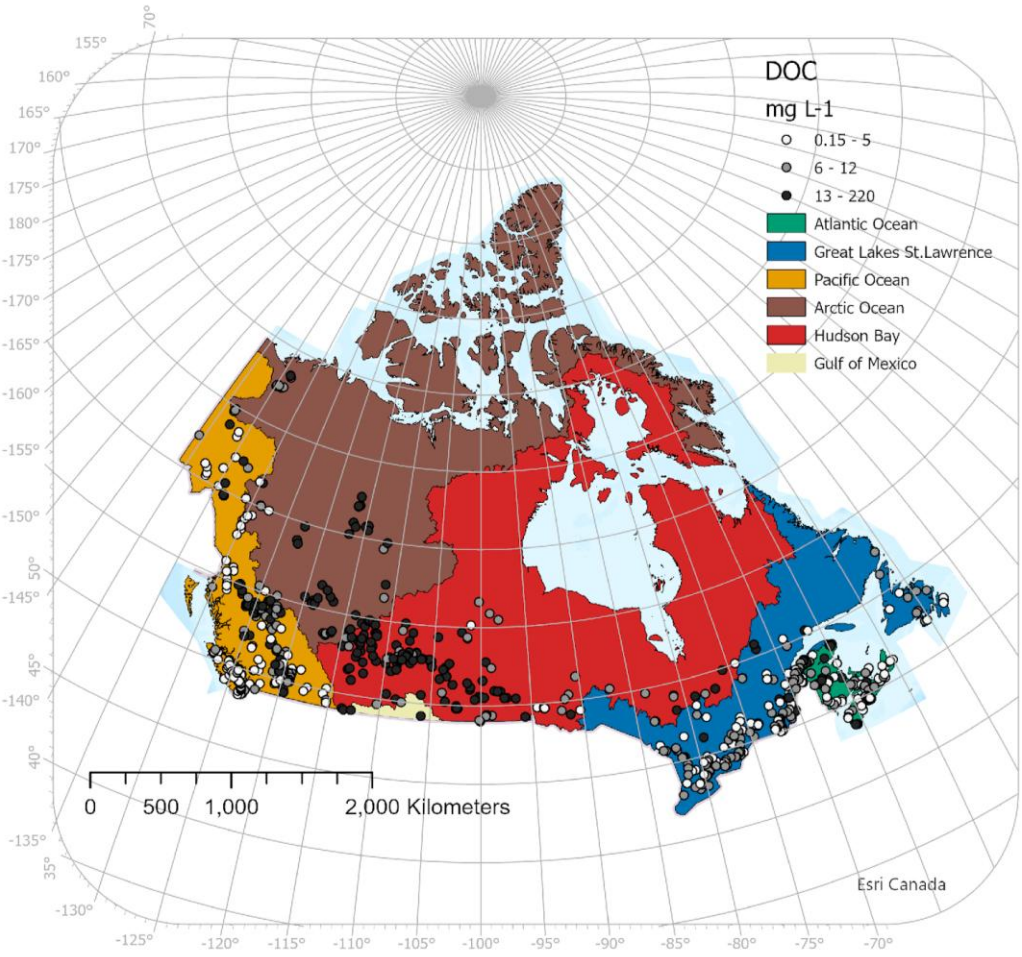

Figure 2.1: Map of sampled lakes by the NSERC Canadian LakePulse Network in Atlantic Ocean, Great lakes St-Lawrence, Pacific Ocean, Hudson Bay, Arctic Ocean continental basins across Canada.

2.3 Materials and method

2.3.1 Study area and sampling

We sampled 664 Canadian lakes for a full set of limnological variables over three years as part of the Natural Sciences and Engineering Research Council of Canada (NSERC), Canadian LakePulse Network (Figure 2.1). Following data cleaning and integration of limnological and DOM composition data, the final dataset included 548 lakes. Lakes were selected based on size, human impact, and landscape

characteristics. To quantify human impact, land use within each watershed was evaluated. Watersheds were delineated using flow direction derived the Canadian Digital Elevation Model (Government of Canada, 2015 Each land use type was assigned an impact value, and the Human Impact Index (0 = natural; 1 = disturbed) was calculated as the mean across all cells in the watershed's land use raster. Land-use patterns were derived from existing remote sensing datasets with a 30-meter resolution. Additional details on lake selection and watershed variables are available in Huot et al. (2019). Lake water samples were collected at the deepest point during July–August of 2017–2019. At each lake, surface water samples were taken using an acid-washed integrated tube sampler from the depth of the euphotic zone, estimated as twice the Secchi depth, to a maximum of 2 meters of water. In situ lake conditions in the water column (i.e., water temperature, conductivity, dissolved oxygen, and pH) were measured using an RBR Maestro multi-parameter water quality meter (RBR Ltd., Ottawa, Canada). Meteorological variables such as air temperature, relative humidity, wind speed, atmospheric pressure, precipitation, and sky condition were measured at each site using a Kestrel 5500 weather meter. Lake morphometric variables, including water retention time (WRT), were obtained from HydroLAKES v1.0 (Messager et al., 2016). WRT was calculated as the ratio of lake volume to the estimated discharge at the lake's outlet, the latter based on regional runoff. To assess the reliability of these modeled estimates, Messager et al. (2016) compared estimated WRT values against observed residence time data from 374 natural lakes, and reported significant positive a correlation ($R^2 = 0.57$), supporting the robustness of HydroLAKES WRT estimates for use in large-scale studies. Samples for CDOM, DOC and chlorophyll, were filtered on the same day in the mobile lab next to the lake and kept in the dark and cold, and together with unfiltered water samples for nutrients and other chemical analyses were sent to the GRIL laboratory, located at the Université du Québec à Montréal, for analysis.

2.3.2 DOM Characterization

Samples for DOM quantification and spectrofluorometric and molecular characterization were filtered using a 0.45 µm syringe filter in the mobile lab and then stored in the dark and cold until laboratory analyses. In the laboratory, CDOM absorbance (230 to 700 nm) was measured on filtered water samples using a Biochrom Ultrospec 3100 Pro spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The specific ultraviolet absorbance was measured ($SUVA_{254}$), by dividing the UV absorbance measured at 254 nm by the DOC concentrations (Cuthbert & del Giorgio, 1992; De Haan, 1993). Fluorescence intensity from lake water samples was measured using a spectrofluorometer (Shimadzu RF5301 PC, Shimadzu, Kyoto, Japan) across excitation wavelengths of 230 – 450 nm (5 nm increments) and

emission wavelengths of 240–600 nm (2 nm increments), and was used to build three-dimensional excitation and emission matrices (EEMs). Corrections for Raman and Rayleigh peaks were carried out using the R package *eemR* (Massicotte, 2016). Fluorescence intensities were normalized to Raman units by dividing the intensity by the Raman area of pure water integrated at an excitation of 350 nm and over an emission range of 380 to 420 nm. Parallel factor analysis (PARAFAC) was used to characterize the DOM fluorescence signal using MATLAB software (MATLAB 7.7.0, The MathWorks, Natick, USA, 2021) and drEEM toolboxes (Murphy et al., 2013; Stedmon & Bro, 2008). The parameters obtained from the PARAFAC model were used to calculate an approximation of the abundance of each component, expressed as Fmax (in R.U.), which corresponds to the maximum fluorescence intensity for a particular sample. Additionally, to investigate the intensity of each PARAFAC component (Ci), the relative fluorescence intensity of each component as a percentage (% Ci) was calculated (%Ci = Ci/ΣCi * 100%). The PARAFAC components were matched to previously reported fluorescence components using the OpenFluor database (Murphy et al., 2014). PARAFAC analysis of EEMs identified five optical DOM components: Component 1 (C1) was a humic-like component (Kothawala et al., 2012); component 2 (C2) was characterized as substances altered by microbial reprocessing (Lapierre & del Giorgio, 2014); component 3 (C3) was characterized as terrestrial humic-like (Kothawala et al., 2014); component 4 (C4) was considered fulvic acid-like (Søndergaard et al., 2003); component 5 (C5) was characterized as protein-like (Lapierre & del Giorgio, 2014). More details on the PARAFAC components can be found in supplementary Figure S2.2. In addition, the humification index (HIX) was calculated from the EEMs, ranging from 0 to 1, where higher values indicate a higher degree of humification (Ohno, 2002). We calculated biological index (BIX) from the EEMs, which corresponded to 0.6 to 0.7 for DOM of low biological components and > 1 for DOM of biological or aquatic bacterial origin (Huguet et al., 2009). Spectral slope ratio (SR) was also calculated as S275–295 divided by S350–400 (Helms et al., 2008).

The molecular composition of DOM was analyzed using Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) on solid-phase extracts obtained from filtered water samples. Extractions was carried out using a modified styrene divinylbenzene polymer-type sorbent (100 mg Bond Elut PPL cartridges, Agilent Technologies) (Dittmar et al., 2008). The vials were kept at -20°C until further analysis. The DOM solid-phase extracts were analyzed using a 7T Bruker SolariX XR (FT-ICR-MS) at the Water Quality Centre, Trent University, Peterborough, Canada. An external calibration was performed each day using electrospray ionization (ESI) and sodium trifluoroacetate (NaTFA) in methanol at a concentration of 0.1 mg mL-1, in negative ion mode. Data acquisition was conducted using the Bruker ftms Control software

(version 2.1.0), covering a mass range spanning from m/z 200 to 1000. The data was exported using the Bruker Compass DataAnalysis software (version 5.0). The samples were injected individually at a flow rate of 120 µL/h to acquire 300 spectra scans. We processed the resulting masses for quality and assigned molecular formulae following established protocols from the ICBM-OCEAN tool, an open platform for DOM mass spectra processing (Merder et al., 2020). Molecular formulas were assigned based on the following elemental constraints: C_{4-100}, H_{4-200}, O_{1-70}, N_{0-4}, S_{0-2}, and P_{0-1}. We cleaned the dataset, samples with sum intensities below 50 were removed, and the intensities of isomers (compounds with the same molecular formula but distinct atom arrangements) were summed. Using the full set of assigned formulas, the relative proportions of compound classes were calculated. In this study, we used four broad chemical compound classes, based on Oxygen/Carbon, Hydrogen/Carbon, and modified aromaticity index (AI_{mod}), which quantifies the degree of aromaticity of an organic compound (Koch and Dittmar, 2016). The classes were "aliphatic" ($H/C \geq 1.5$), "Low Oxygen (O) unsaturated" ($H/C < 1.5$, $AI_{mod} < 0.5$, $O/C < 0.5$), "High Oxygen (O) unsaturated" ($H/C < 1.5$, $AI_{mod} < 0.5$, $O/C \geq 0.5$), and "aromatics" ($0.5 < AI_{mod} < 0.67$). These classes, previously used in other published studies, represent broad compositional groups highlighting key chemical differences between the samples and demonstrating inter-instrument comparability (Hawkes et al., 2020). The DOM molecules were classified into chemical compound classes using the ftms analysis package, adapted from Bailey et al (2017). The classification involved assessing the number and type of molecules in the different classes in each sample, irrespective of the molecule peak intensity.

2.3.3 Stable isotope of water and deuterium excess

Samples for δ^2H and oxygen $\delta^{18}O_2$ analyses were collected from the surface waters at the deepest point in each lake and preserved in 30 mL HDPE bottles, and kept at 4℃ in the dark until further analyses. We analyzed δ^2H and oxygen $\delta^{18}O_2$ in water using an LGR (Los Gatos Research) model T-LWIA-45-EP Off-Axis Integrated Cavity Output Spectroscopy (OA-ICOS) at the GEOTOP laboratory, located at the Université du Québec à Montréal. Deuterium excess (d-excess) values were calculated as follows: d-excess = $\delta^2H - 8 * \delta^{18}O_2$ (Dansgaard, 1964). d-excess represents a departure from the local meteoric signal and as such can be used as an index of the degree of evaporation in a given water sample.

2.3.4 Statistical analysis

Data analysis, visualization, and statistical computations were conducted using R v. 4.1.0 (R Core Team 2021). We conducted checks on the normality of environmental data. When necessary, we transformed the data to log10; for d-excess, a log10 + X transformation was applied, and for fractional watershed

variables, a square root approach was used. We normalized the DOM molecular formulae intensities within each sample to the total intensity of the sample. We aggregated the data into six continental ocean drainage basins in order to explore regional patterns of DOM concentration and composition: Atlantic Ocean, Great Lakes–St. Lawrence, Gulf of Mexico, Hudson Bay, Arctic Ocean and Pacific Ocean. To balance continental basin observational group sizes, we merged lakes located in the Gulf of Mexico (n = 2) to Hudson Bay basin. These patterns were explored by using boxplots, and statistical differences among regions were assessed using one-way ANOVA and Tukey's HSD post hoc tests. Principal components analysis (PCA) was used to explore how different optical and chemical components of DOM relate to each other using *ggfortify* package's *precomp*() function. We then used the *kmeans*() function from the *cluster* package to delineate lake clusters based on DOM composition and to identify regional-scale geographic patterns in these groupings. We used silhouette coefficients to determine the appropriate number of clusters. We applied partial least squares regression to investigate how different lake, catchment, and climate variables (X predictors) relate to DOM concentration and composition (Y responses: DOM concentration, molecular classes, and optical components and indices). PLS regression is particularly suitable for ecological datasets with collinear predictors and non-normal distributions, as it is relatively robust to correlation among X variables, deviations from normality, and missing data. The results of the PLS analyses were visualized using loadings plots. In these plots, the first component typically explains more variance than the second, rendering it more influential for interpretation. Variables positioned close together are positively correlated, those at opposite ends of the plot are negatively correlated, and variables near the plot center show weak relationships with others. To complement the PLS analysis and to evaluate the combined effects of d-excess and TP on the composition of DOM as the most important variables identified by the PLS, we developed multiple regression models as a function of d-excess and TP, with basin as a qualitative moderating factor. The interaction term between TP and basin allowed us to explore the effect of TP on DOM composition across different regions. This tested whether regional differences modulate the relationship between TP and DOM composition, while accounting for hydrological influences. Furthermore, as variable importance in projection (VIP) scores from the PLS analysis identified d-excess as the dominant driver of both DOC concentration and DOM composition, and because d-excess provides unique insights into terrestrial–aquatic connectivity that is not captured by other variables such as TP, we focused on d-excess thereafter. We assessed regional differences in the relationship between different molecular classes of DOM and d-excess by conducting an analysis of covariance (ANCOVA) using the stats package's *lm*() function. The molecular classes were set as the dependent variable, region as the categorical predictor, and d-excess as independent variable. We first

tested models with different slopes, implying that the relationship between hydrology and DOM classes may vary across regions. This enabled us to assess potential differences among continental basins while considering the effects of environmental variables on DOM molecular composition. If these interactions were not significant, we continued with models that had fixed slopes but varying intercepts, which assumed varying baselines of DOM across regions, and further assessed how and why these baselines vary.

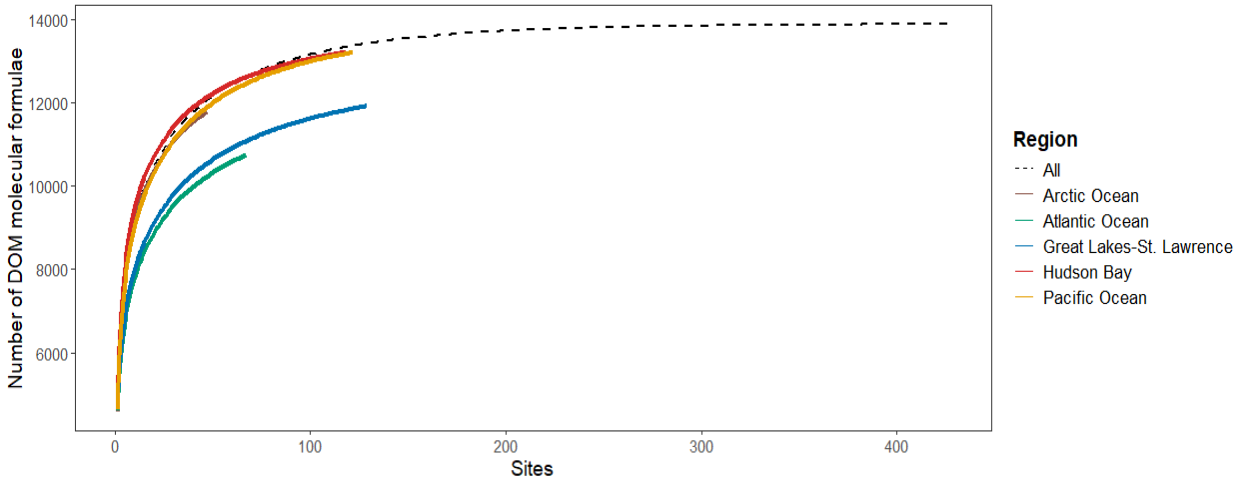

Figure 2.2: Rarefaction curves of DOM molecular richness for all samples together and by continental basin.

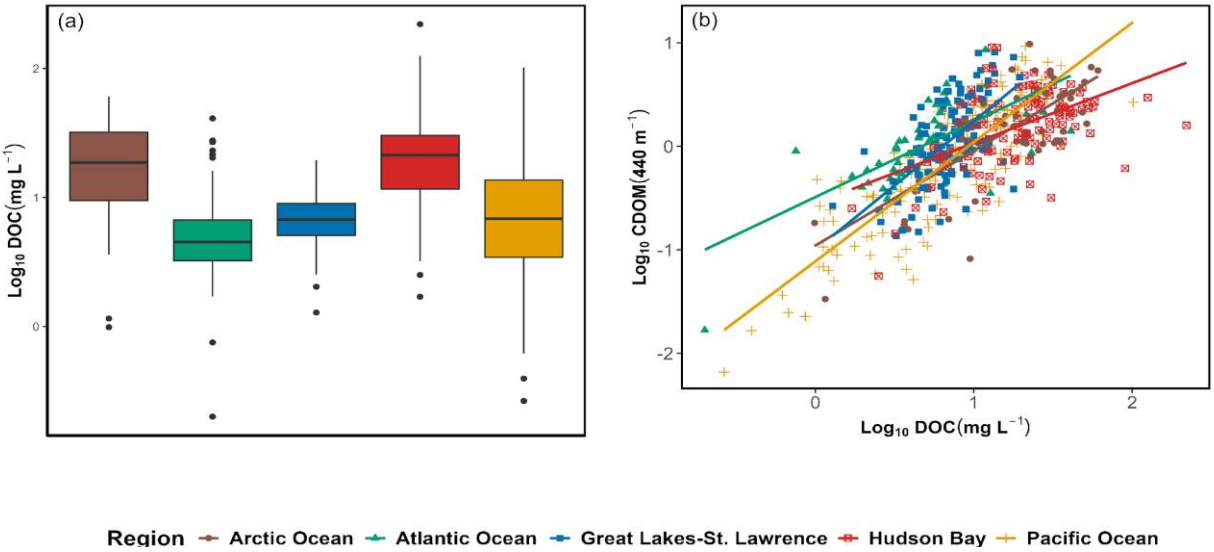

Figure 2.3: Boxplot showing the DOC concentration (log_{10}) across five regions (a). DOC concentration (log_{10}) vs. the log_{10} CDOM absorbance coefficient at 440 nm across five continental basins (b). Post hoc Tukey's HSD results: Arctic Ocean had significantly higher DOC concentrations compared to the Atlantic Ocean, Great Lakes-St. Lawrence, and Pacific Ocean basins. Hudson Bay also had significantly higher DOC concentrations than the Atlantic Ocean, Great Lakes-St. Lawrence, and Pacific Ocean basins. The Pacific Ocean had significantly higher DOC concentrations than the Atlantic Ocean and Great Lakes-St. Lawrence, but lower than Hudson Bay. Great Lakes-St. Lawrence had slightly higher DOC concentrations than the Atlantic Ocean, though this difference was not significant.

2.4 Results

2.4.1 Environmental variables and DOM analysis

Watersheds were distributed among the five major Canadian basins, with sizes ranging from 0.2 to 37,459 km². The lakes also varied greatly in size and morphometry, from 0.007 to 266 km² and from 1 to 64 meters in maximum depth (Zmax). DOC concentration varied widely, from 1 to 220 mg L^{-1}, with CDOM ranging from 0.006 to 9.7 m^{-1}. The hydrologic conditions were very heterogeneous across these lakes, with d-excess ranging from -37.1 to 21.1 permil (‰), and estimated WRT from 0.1 to 9776 days; there was an overall significant but rather weak negative relationship between the estimated WRT and the measured d-excess in these lakes (Supplementary Figure S2.1). This pattern suggests that, while longer WRT enhances evaporative enrichment, leading to lower (more negative) d-excess, other hydrological factors also influence the isotopic signature. Specifically, d-excess not only reflects local evaporative processes linked to WRT, but also integrates the isotopic history of upstream waters, which include the origin of the water, its transit time within the watershed and the various pathways it has followed before arriving at a particular lake, and this is what renders this metric so powerful from a predictive perspective. Detailed information regarding these, and other environmental and watershed characteristics can be found in Supplementary Table S2.1.

We identified a total of 13,440 distinct molecular formulae across all lakes. The average number of formulae varied between continental basins (Figure 2.2), but in all basins we reached a plateau, so we captured the extant molecular diversity. Based on molecular characteristics, compounds were grouped into four major categories: high O unsaturated (33.7%), low O unsaturated (32.8%), aromatic (22.5%), and aliphatic (11%). These percentages represent the proportion of each molecular class relative to the total formulae pool.

2.4.2 Patterns in lake water isotopes across Canada

The isotopic distinctions between basins across Canada is well reflected in $d^{18}O$-$d^{2}H$ relationship (Supplementary Figure S2.3). Our results showed that $d^{18}O$ values were highly correlated with $d^{2}H$ signatures and there were distinguishable clusters based on basins. The Great lakes St. Lawrence and Atlantic Ocean basins were clustering near the Canadian water meteoric line (CWML) suggesting a stronger influence of precipitation and minimal evaporative enrichment. In contrast, a subset of lakes from the

Pacific Ocean, Hudson Bay, and Arctic Ocean basins deviated more substantially from the CWML, indicating greater evaporative influence.

2.4.3 Patterns in DOC concentration and CDOM

Figure 2.3a shows that the highest mean DOC concentration values were observed in the Hudson Bay and Arctic Ocean basins, and the lowest were observed in the Atlantic Ocean basin (22.1, 21.3, and 7.3 mg L^{-1}, respectively). The overall ANOVA for DOC concentration was statistically significant (ANOVA: $F=64.45$, $p<0.0001$), indicating significant differences in average DOC concentrations among basins. CDOM varied by three orders of magnitude across lakes, from 0.0065 to 9.71 m^{-1} and there was an overall positive relationship between DOC and CDOM (Figure 2.3b), which differed between basins, reflected in differences in both the slope and intercept of their respective regression models ($p<0.0001$, Supplementary Table S3.2). There were up to 2 orders of magnitude range in CDOM for any given DOC concentration, especially at the high DOC concentration ranges, suggesting major qualitative or compositional differences among sites and regions. Lakes in the Atlantic Ocean basin ($p<0.0001$) tended to have high CDOM (and a higher intercept of the CDOM - DOM relationship), whereas the Pacific Ocean basin ($p<0.0018$) tended to have on average lower CDOM, and the lowest intercept of the CDOM - DOC relationship (Supplementary Table S3.2).

2.4.4 Patterns in optical and molecular DOM composition across lakes

We conducted a PCA analysis on both molecular classes and optical components (here expressed in relative contribution to total fluorescence), followed by a Kmeans clustering analysis, which resulted in three major clusters of lakes (Figure 2.4). The first two axes together explained 61% of the variability in DOM composition. Figure 2.4 shows that there were strong links between the molecular composition and the optical characteristics of DOM in lakes across Canada. The dominance of PARAFAC component %C1 was strongly related to the Aromatics class. Both were positively related to SUVA, and strongly negatively correlated to the bulk H/C of the sample (determined from FT-ICR-MS) and to the SR and BIX indices; this pool of DOM overwhelmingly dominates PC1. Component %C5 was aligned with the Aliphatic class, and a mirror image of the humification index (HIX). Component %C3 was strongly coupled to the High.O.unsat class, positively to the bulk O/C of the sample (from the FT-ICR-MS), and secondarily to HIX. In the exact opposite direction of PC2, and weighing heavily on this axis, Component %C4 covaried strongly with the Low.O.unsat class. The optical component that was least related to the molecular composition was %C2,

perhaps because this optical category would require a finer classification of molecular properties to reveal any significant link.

The lakes tended to aggregate and form clusters based on their similarities in terms of the dominant pools of DOM (Figure 2.4). There appear to be three main clusters in terms of DOM composition. These clusters are primarily distinguished by differences in Aromatic, Aliphatic and High.O.unsat. Moreover, the clusters show some regional patterns, suggesting that geographic factors may influence DOM composition. Cluster 1 was composed predominantly of lakes in the Hudson Bay and Arctic Ocean basins, whose DOM was characterized by proportionately elevated %C3, High.O.unsat and O/C, and secondarily, high BIX and SR. Cluster 2 is very geographically heterogenous with lakes located in all basins, and DOM dominated by component %C5 and the Aliphatic class, and secondarily by component %C4 and Low.O.unsat, and characterized by high BIX and SR and low HIX. Finally, cluster 3 was populated by lakes mostly located in the Great lakes -St Lawrence and Pacific Ocean basins, with DOM dominated by high component %C1 and class Aromatics, and characterized by high SUVA, low bulk H/C and low BIX and SR indices, suggesting terrestrially dominated but less processed DOM overall.

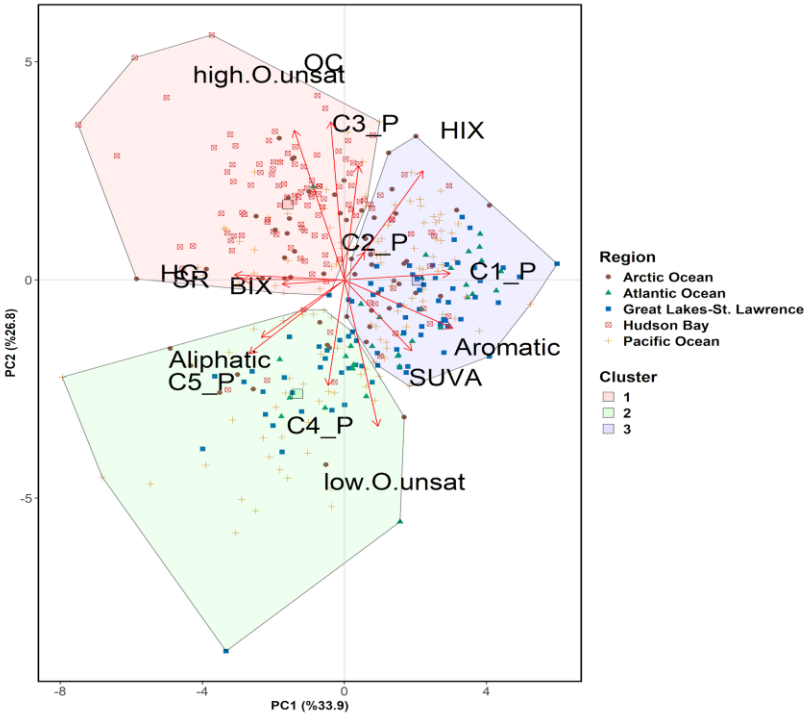

Figure 2.4: Principal component analysis (PCA) showing the relationship between different optical components, indices and molecular classes across studied lakes.

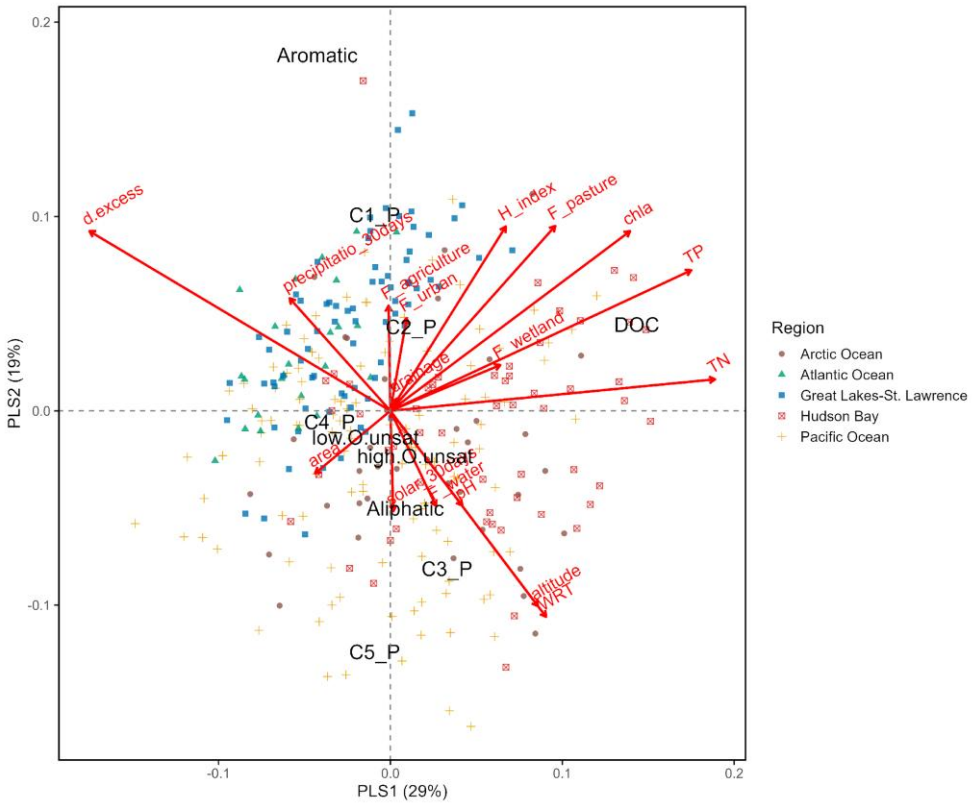

Figure 2.5: Partial least square regression (PLS) showing the environmental drivers (shown in red) of DOC concentration and different molecular classes and PARAFAC components (shown in black). Abbreviations: DOC – dissolved organic carbon, TP – total phosphorus, TN – total nitrogen, chla - chlorophyll a, , area – lake area, WRT – water residence time, D-excess – deuterium excess, F-water – fraction of water, H-index – human impact index, F-urban – fraction of urban, F-pasture – fraction of pasture, F-agriculture – fraction of agriculture, precipitation_30days – 30 days period precipitation, solar_30days – 30 days period solar radiation, pH – pH, altitude – altitude, drainage – drainage ratio, Aromatics – Aromatic, Aliphatic – Aliphatic, C1_P – %C1, C2_P – %C2, C3_P – %C3, C4_P – %C4, C5_P – %C5, high.O.unsat – High.oxygen.unsaturated, low.O.unsat – Low.oxygen.unsaturated.

2.4.5 Drivers of DOC concentration and DOM composition

To evaluate how DOC concentration and its various molecular and optical components are influenced by land-use, hydrology and climate, and in-lake characteristics, we performed a PLS analysis, which accounted for 48% of the overall variation (Figure 2.5). The PLS analysis shows that DOC concentration was strongly positively related with TP and WRT, and strongly negatively related with d-excess. Neither lake area nor the drainage ratio were effective predictors of DOC concentration. In contrast, watershed properties, including the fraction of wetlands, pasture, and surface water, were associated with DOC (Figure 2.5). The

PLS model revealed that d-excess was the most significant variable influencing DOM composition, with positive relationships with classes Aromatics and Low.O.unsat, and components %C1, %C2, and %C4. Conversely, classes Aliphatic and High.O.unsat and components %C3, and %C5 were negatively associated with d-excess and positively linked to WRT. TP was also associated to the optical components and to the molecular classes, with %C1, %C2, and %C3 increasing, and %C4 and %C5 declining as a function of TP, whereas Aromatics and High.O.unsat classes increased, and Aliphatic and Low.O.unsat classes decreased as a function of TP. Additionally, class Aromatics and component %C1 were negatively correlated with the human impact index and the fractions of wetland and pasture (Figure 2.5). Higher DOC and dominance of Aliphatic class and %C1 component of DOM in TP-rich, highly evaporated water suggests a major effect of internal production for lakes in Hudson Bay and Arctic Ocean basins.

We calculated variable importance in projection (VIP) scores from PLS analysis (Supplementary figure S2.6). The VIP scores confirmed that d-excess and TP were the two most influential predictors across all models, reflecting the influence of hydrology (as indicated by d-excess) and trophic status (as indicated by TP). Variables such as TN and *Chla* are closely associated with trophic status and strongly covary with TP. Based on these results, we developed multiple regression models of DOC, molecular and optical properties based on d-excess, TP, and basin as a categorical variable (Supplementary Table S2.3), and we tested for regional differences in the relationship between TP and DOM concentration and composition, while accounting for d-excess influences. These multiple regression models assumed the effect of d-excess on DOM concentration and composition is constant across all regions. The strength of the overall relationships, and the sign of the individual coefficients varied between models, and also the significance of the regional effect varied greatly among models and between variables. DOC concentration was strongly negatively related to d-excess ($p < 0.001$) and positively related to TP ($p < 0.001$), and there were significant regional interactions (TP * basin) in this model. All models for molecular classes were highly significant for both d-excess and TP, with Aromatics class having positive coefficients with d-excess and TP, Aliphatic class having the inverse pattern, High.O.unsat class having a negative coefficient with d-excess and positive with TP, and Low.O.unsat class having the inverse pattern. In all cases, there were significant regional interactions, suggesting that relationships between TP and DOM composition may be modulated regionally (Supplementary Table S2.3). Models for optical components were significant for both d-excess and TP as well, with %C1, %C2 and %C3 having positive coefficients with both d-excess and TP, and %C4 and %C5 had negative coefficients with both d-excess and TP. In all cases, there were significant regional

interactions, implying that relationships between TP and DOM optical components were strongly influenced by regional factors (Supplementary Table S2.3).

Table 2.1: Analysis of variance table for different models.

Model	F value	Pr(>F)
Aromatics vs d-excess		
Log10 d-excess	288.63	<0.001***
Region	2.06	0.083.
d-excess*Region	2.58	0.03*
Aliphatic vs d-excess		
Log10 d-excess	19.39	0.001***
Region	3.30	0.01*
d-excess*Region	6.92	0.001***
High.O.unsat vs d-excess		
Log10 d-excess	444.40	<0.001***
Region	18.19	0.001***
d-excess*Region	4.62	0.0011**
Low.O.unsat vs d-excess		
Log10 d-excess	165.98	< 0.001***
Region	27.2	< 0.001***
d-excess*Region	3.45	0.0084**

Signifiant : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2.4.6 Regional differences in the relationship between Dom classes and hydrology

Since hydrology and the extent of evaporation, as reflected in water stable isotopes, appear to play a critical role in shaping DOM concentrations and composition across Canadian lakes, we developed individual empirical regression models for Aromatic, Aliphatic, High.O.unsat, and Low.O.unsat classes with d-excess as independent variable. We included the continental basins (regions) as a categorical variable in all the models, but did not include TP since it was difficult to assign regional or fixed values for this variable. We developed parallel models for DOC concentration and PARAFAC components (Supplementary Figure S2.4 and S2.5). The relationships between d-excess and Aromatics, Aliphatics, High.O.unsat and Low.O.unsat varied greatly in both slopes and intercepts across the different basins (Figure 2.6 and Table S2.4), suggesting that the direction and strength of these relationships differ among DOM pools and geographically. Our results show that as water becomes more evaporated (lower d-excess, longer WRT) within the lake and aquatic network, certain molecular DOM classes such as Aromatics and Low.O.unsat and the associated PARAFAC component C1 tend to decline in absolute terms (Figure 2.6, Supplementary Figure S2.4), whereas other DOM pools, such as Aliphatic, High.O.unsat and the associated C5 tend to become enriched (Figure 2.6, Supplementary Figure S2.4). A similar trend of becoming more enriched was

observed in DOC concentration (Figure S2.5). It is also apparent that for any given d-excess, lakes in different regions may have different absolute DOM compositions, and this suggests that the initial composition of freshly loaded DOM to the aquatic network may vary across regions. Additionally, d-excess may act as a proxy for the extent of retention within the network and therefore of degradation, in terms of the hydrologically mediated departure from the original source material, and this appears to be a major driver of DOM composition in lakes across Canada.

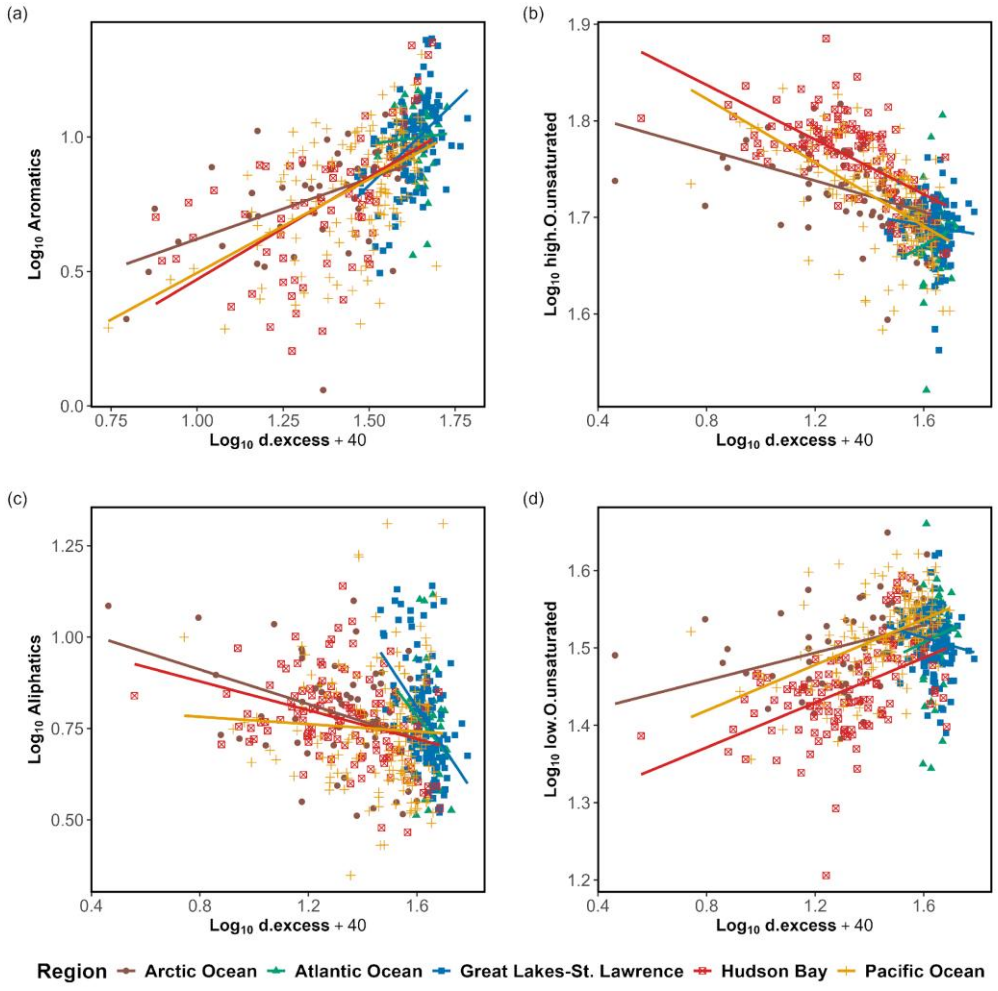

Figure 2.6: Relationships between molecular classes (Aromatic, Aliphatic, High.O.unsat, Low.O.unsat) and d-excess across different regions.

2.4.7 Reconstructing the composition of source and processed DOM

We further used these empirical models that relate DOM composition to d-excess (and therefore to the degree of evaporation and potentially of network water age or processing) to extrapolate DOM composition that might be associated to the least evaporated water and therefore to freshest or least processed pools of DOM within the aquatic network. We did this by estimating the composition of the DOM at a fixed valued of d-excess = 20, which corresponds to the least evaporated water within our data base and likely represents the most recent precipitation or groundwater inputs, and we applied this fixed value to the models for all the basins and the four molecular components. This allowed us to reconstruct the overall DOM composition associated to recent water inputs to lakes in each basin. Likewise, we estimated the expected DOM composition at the other extreme of the hydrologic continuum, associated to the most evaporated (and likely aged water), and therefore likely having undergone the most processing, and for this we applied the lowest d-excess values (most evaporated water) observed in each basin (not an assumed, fixed d-excess value as for the source DOM estimation). We finally assessed if the regional reconstructed intercepts from the models described above were related to regional properties, including average regional topography, soil properties, and climate (Figure 2.8, Supplementary Table S2.4). Figure 2.7 shows the regional variability of the four molecular classes in this reconstructed source DOM, showing differences in their distribution across the five major basins. There were relatively minor differences in the average reconstructed composition of the source DOM between regions (Figure 2.7a). High.O.unsat emerges as the most abundant compound class in the source DOM of all basins, followed by Low.O.unsat, Aromatics and Aliphatics, the latter having on average a low overall intensity in the source material in all regions. The largest differences in the composition of the source material observed between regions corresponds to a substantially higher contribution of Aromatics in the Great Lakes, a lower contribution of Low.O.unsat in the Great Lakes and in the Hudson Bay, and a low contribution of High.O.unsat in the Pacific Ocean.

We further reconstructed the composition of DOM that had been processed within the network at the maximum d-excess (i.e. the most evaporated water) observed in each region, again using the empirical equations that relate compound classes to d-excess (Figure 2.7 and Supplementary Table S2.4). The composition of this processed DOM (Figure 2.7b) diverges greatly from that of the source material within the same basin (Figure 2.7a). High.O.unsat and Aliphatic classes substantially increase in the processed DOM relative to the source DOM, whereas Aromatics and Low.O.unsat classes consistently decline along the same hydrologic gradient (Figure 2.7b). Although this pattern was consistent across basins, there were

nevertheless substantial differences in the composition of processed DOM among basins. Low.O.unsat and Aromatics had the steepest declines in the Arctic Ocean, Hudson Bay and Pacific Ocean basins, whereas the largest increases in High.O.usat compounds occurred in the Arctic Ocean, Hudson Bay, and Pacific Ocean. Aliphatics almost doubled in the Arctic Ocean and Great Lakes basins. We further assessed if the reconstructed composition of the source DOM (Figure 2.7a) derived from the models described above (Figure 2.6, Supplementary Table S2.4) was related to regional watershed features, including average regional topography, soil properties, and climate. The reconstructed average source DOM intensity of Aliphatics class was positively related to the average regional altitude (Figure 2.8a), whereas the reconstructed average source DOM intensity of the Aromatics class was positively related to the average regional soil organic content (SOC) (Figure 2.8b). The contribution of Low.O.unsat and High.O.unsat compounds to the source DOM was not related to any watershed or regional variable that we explored.

2.5 Discussion

Here we assessed DOM concentration and composition in lakes across five major continental watersheds in Canada. Although there was an overall relationship between DOC and CDOM across all the study lakes, regional differences led to significant variations in average CDOM levels for any given DOC concentration among the different basins. We have shown that these variations in the CDOM/DOC relationship are linked to changes in the optical properties of DOC across lakes and regions, as revealed by EEMs analyses. Such regional variations in optical properties have been observed before, and suggest major differences in lake DOM composition both within and among regions, and we have shown that there is a strong link between specific optical components and broad chemical categories derived from the molecular analysis of DOM. We were able to further discern the most important factors explaining the variability in DOM concentration and composition across Canadian lakes, which included proxies of lake trophic status (TP) as well as variables related to regional and lake hydrology. In particular, d-excess, a proxy for water evaporation and thus for water residence time and water history within the aquatic network, emerged as the single strongest variable shaping the composition of DOM in lakes at a continental scale, highlighting the importance of hydrology in regulating both total DOM and its optical and chemical composition. We showed that the relationships between individual DOM molecular classes and d-excess varied significantly in their slopes and intercepts among basins, resulting in regional differences in DOM composition along hydrologic gradients. We have further shown that there are subtle differences in the composition of the reconstructed source DOM among regions, linked in part to regional landscape and topographic features. There were even larger differences in the network processing of this DOM along the hydrologic continuum

among basins, leading to further differences in the average DOM composition among regions. The source DOM and the extent of its processing along the aquatic network together shape the DOM composition in lakes across Canada, yet our results suggest that the latter plays the leading role in determining cross regional patterns in DOM composition.

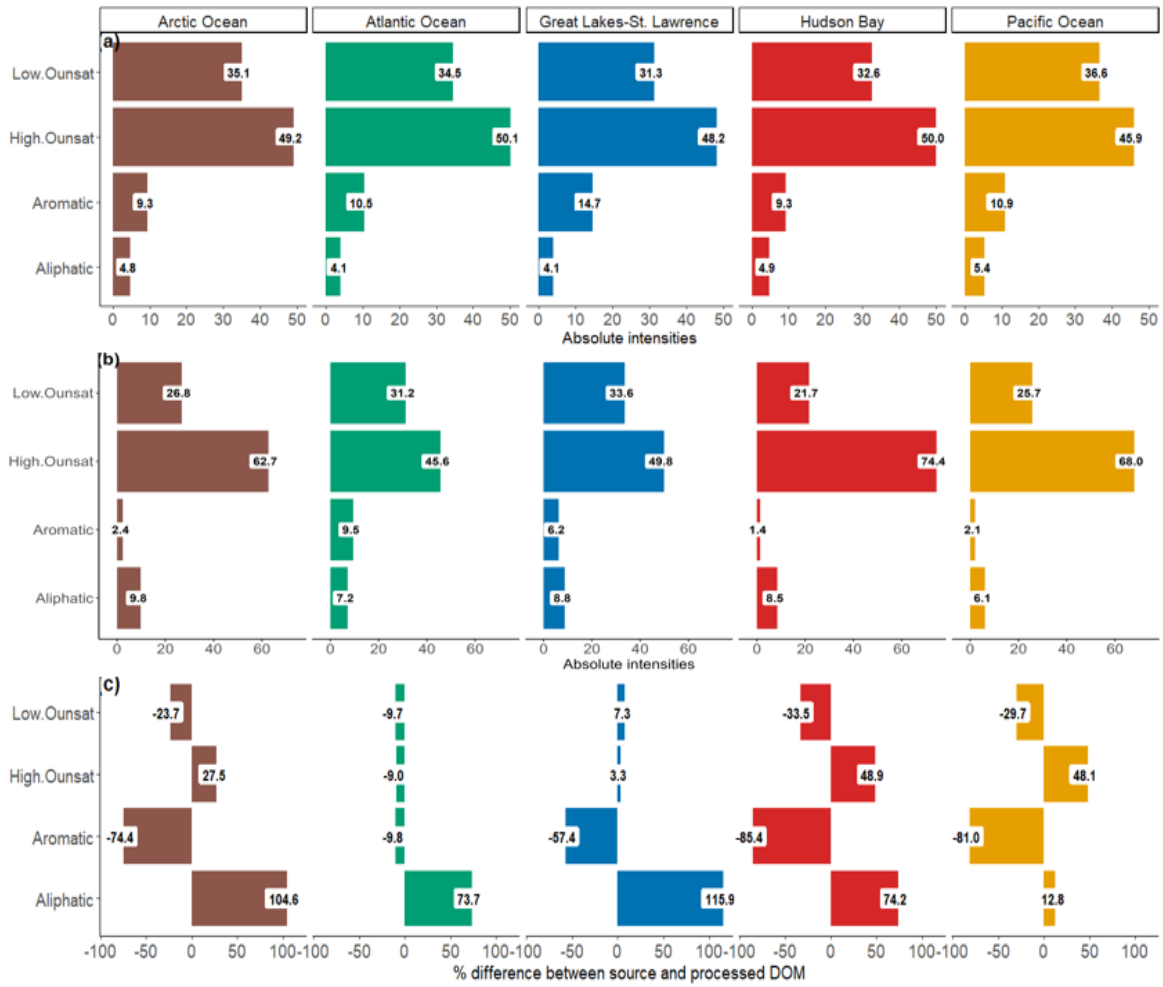

Figure 2.7: Estimated total signal intensity of each molecular class for DOM associated to newly loaded water from fresh precipitation (assuming a d-excess for all regions = 20) (a) and Total signal intensity of each molecular class for DOM that had been processed within the network at the maximum d-excess observed in each region (b). Difference in the source (panel (a)), and processed DOM (panel (b)) in each region (c).

2.5.1 Links between optical and molecular properties of DOM

DOC concentration and CDOM (absorbance at 440 nm) were strongly positively correlated across our study lakes (Figure 2.3). Previous research has consistently demonstrated a relationship between CDOM and DOC concentration (Kurek et al., 2024; Lapierre et al., 2013; Mann et al., 2016; Spencer et al., 2012; Stedmon et al., 2011). This relationship has also been shown to vary spatially and also in time (Johnston et al., 2020; Kurek et al., 2023; Liu et al., 2021) and this variation in the CDOM/DOC relationship has major implications on the development of models and robust and scalable algorithms for remote sensing of DOC, among others (Liu et al. 2021; Deutsch et al. 2022). The processes underlying the variability in the CDOM/DOC relationship are not well understood, and here we have shown that these variations are likely associated to shifts in both optical and molecular properties of the DOM that occur at continental scales. The broad spatial scope of our study facilitated the identification of associations between specific optical components and particular molecular classes within the DOM, pattern that has also been reported previously (Johnston et al., 2020; Orlova et al., 2024; Kurek et al., 2024; Stubbins et al., 2014). We have shown that PARAFAC component %C1 and SUVA were strongly related to the Aromatics class, suggesting that these indices provide similar information and likely comprises high molecular weight humics. Component %C5 was aligned with the Aliphatic class, and a mirror image of the humification index (HIX), suggesting that this pool of DOM is more prevalent in samples with less accumulation of humic degradation products. Component %C3 was strongly coupled to the High.O.unsat class, positively to the bulk O/C of the sample (from the FT-ICR-MS), and secondarily to HIX, suggesting that these humic pools of DOM are involved in extensive degradation and transformation processes.

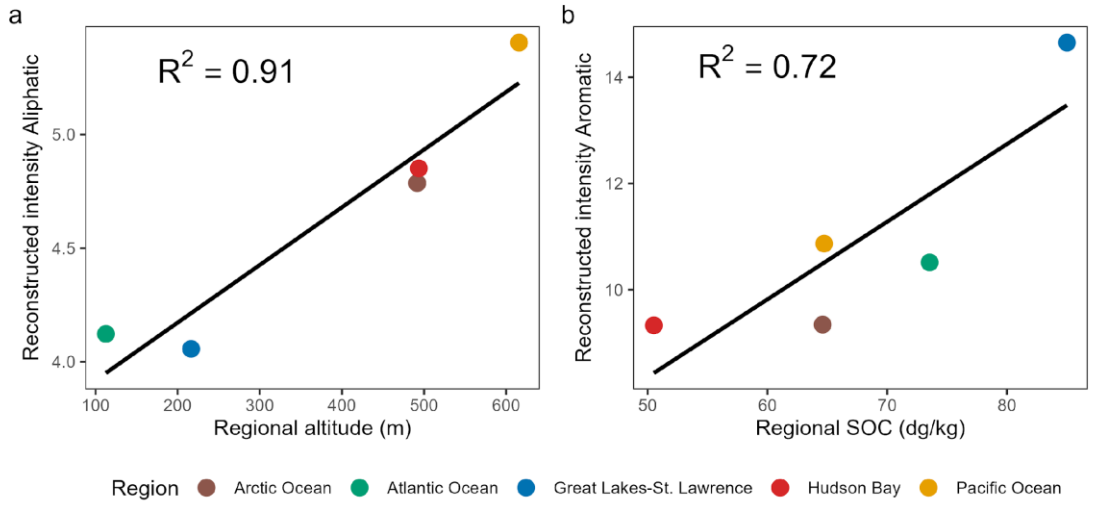

Figure 2.8: The average regional intensities for Aliphatics and Aromatics for the reconstructed source DOM as a function of the average regional altitude (a) and soil organic carbon (b) across different regions.

2.5.2 Regional differences in the network processing and sources of DOM

The results obtained from the empirical regression models have revealed significant variation in the region-specific intercepts and slopes of DOM composition models as a function of d-excess (Supplementary Table S2.4). The basins also differed greatly in the total range of d-excess found across their lakes, suggesting major differences in network hydrology and residence time of the studied lakes. This is evident in Figure 2.6, suggesting that the extent of evaporation and the associated water residence time varies greatly between basins, and therefore the total exposure of DOM to hydrologic processing, and the intensity of this processing (as reflected in the slopes) vary among regions. The varying extent of evaporation in the different basins is reflected in the contrasting DOM composition profiles that we observed at the d-excess endpoint of each basin, as was shown in Figure 2.7b, presumably the result of different overall exposure of the DOM to processing. Lakes in the Hudson Bay and Arctic Ocean basins have a wider range in d-excess (Figure 2.6) and longer WRT (Supplementary Figure S2.1), and tend to have on average higher DOM concentration that appears to be more processed, as evidenced by a decline in the absolute amounts of Aromatics and Low.O.unsat classes relative to other regions (Figure 2.7b), in agreement with previous studies (Kurek et al., 2023; Osburn et al., 2011). In contrast, lakes in the Great Lakes St. Lawrence and Atlantic Ocean basins have a relatively narrow d-excess range, and appear to be more connected to their surrounding watersheds and have on average shorter WRT (Supplementary Figure S2.1), which leads to less evaporative signals, and to DOM that appears to have undergone less processing in these regions (Figure 2.7b), with higher concentrations of Aromatics and Low.O.unsat photolabile compounds, also in agreement with previous studies (Creed et al., 2008; Mangal et al., 2017). It is clear that the regional differences that we observed in average lake DOM composition are directly related to differences in the extent of hydrologic processing, which are likely driven by a combination of climate, topography and the architecture of the aquatic network.

There was a remarkable difference between the reconstructed composition of source DOM and the composition of lake DOM at the extreme of the evaporation gradient at each region, but these differences varied by region. In all regions, there was a net removal of Aromatics, in some regions by as much as 85%. In most regions (except in Great Lakes St. Lawrence), there was also a net removal of Low.O.unsat relative to the source material by as much as 37%. In contrast, in all regions there was a net addition of Aliphatics (some regions by over 100%) and High.O.unsat (some regions by over 50%) during transit within the aquatic network (Figure 2.7c). Prior studies have also shown selective removal and addition of certain DOM pools could occur in lakes (Osburn et al., 2011). In this study we have shown that the extent of these net removals and additions varied regionally and were a function of the extent of the regional evaporation gradient, with the largest removals and additions in basins with the largest evaporation gradients. Our results confirm that there is net production of Aliphatics and High.O.unsat along a gradient of residence time and water evaporation in Northern aquatic networks, but the processes that generate these compounds are likely very different: The former likely represents organic matter of aquatic origin that results from various aquatic metabolic processes (Hosen et al., 2021), whereas the latter represents terrestrially-derived organic matter that has undergone significant processing and that tends to persist and accumulate in the DOM pool (Behnke et al., 2021; Kurek et al., 2024), so the process that lead to accumulation of these two pools along a gradient of evaporation are fundamentally different. We have shown that High.O.unsat and optical component %C3 are strongly correlated, confirming that this PARAFAC component represents a highly photodegraded, persistent humic pool. Many studies have assumed that PARAFAC omponent 5 (protein-like or tryptophan-like) represents colored DOM of autochthonous origin, but few have actually demonstrated this assumption (Kurek et al. 2024). Here we have shown that %C5 is strongly correlated to Aliphatics, so that this optical component is indeed coherent with this chemical group, and we have further shown that although the source DOM, likely soils, contains Aliphatics, this class is greatly enriched during transit within the aquatic network. We have also confirmed that %C1 is strongly linked to Aromatics, a terrestrially derived pool that is most strongly linked to hydrology and evaporation (and therefore to exposure), and likely the most photoreactive DOM pool.

2.5.3 Main drivers of DOM concentration and composition

Our results suggest that at a continental scale, hydrology, including lake water retention time and the network water history as reflected in the isotopic composition of lake water, interacts with lake trophic status, as reflected in the lake TP concentration (and the watershed variables that determine the latter), to shape not only DOC concentration, but also the optical and molecular composition of DOM, and that

these relationships are further modulated regionally. We showed that TP concentration as a proxy for a trophic status of the lake was a strong predictor of DOM concentration and composition (Figure 2.5). It is important to acknowledge that TP is a fundamental component of natural OM and can co-vary with DOC across lakes, and this may explain in part the observed relationship between DOM and TP. We should point out that in this Canada wide data set, there was a poor relationship between TP and DOC (R^2 = 24), so there was a large degree of decoupling between the two. In addition, our analyses clearly point to the fact that the effect of TP is associated to lake trophic status, since TN and *Chla* also emerged as important predictors alongside TP. We conclude that the effect of TP on DOM that we observe is mediated mostly by the influence that TP has on lake trophic status and productivity, and is not a simple reflection of the potential covariation of TP and DOC across lakes. The link between lake trophic status and DOM has been previously reported in other studies (Wang et al., 2024; Wen et al., 2022), highlighting that higher TP levels are associated with increased contributions of autochthonous DOM, which tends to be more labile and less aromatic. Our results suggest that TP has an overall positive influence on DOM concentration, likely exerted through its role in modulating lake primary production and overall ecosystem metabolism (Holgerson et al., 2022). Our results further suggest that TP influences DOC concentration by enhancing specific DOM components. In particular, we found that there was a positive relationship between TP and both the Aliphatics and the High.O.unsaturated components, which several previous studies have linked to autochthonously produced DOC (Liu et al., 2022), whereas DOM components that are mostly terrestrially derived did not relate to TP. This points to a key role of internal lake processes in generating specific pools within the DOM, driven in large part by nutrients. In this regard, although for simplicity we focused on TP here, our results suggest that TN also plays a major role is shaping DOM dynamics. In a parallel study, we showed that these DOM components that are linked to TP are indeed associated to primary production in lakes and added to the DOM pool though internal lake processes along the aquatic continuum. In turn, these Aliphatic-rich compounds and their associated optical components appear to fuel ecosystem respiration, and in turn, that respiration depletes these compounds, suggesting that these DOM components are key substrates mediating the relationship between primary production and respiration, particularly in eutrophic lakes, whereas in oligotrophic lakes, terrestrially derived DOM fractions often play a larger role in supporting ecosystem respiration. This suggests that TP may indirectly shape DOM molecular characteristics by influencing the balance between autochthonous and allochthonous DOM sources, as well as the extent of microbial reworking of DOM. It is interesting to note that the effect of TP on DOM varied among basins, such that comparable TP had very different effects on

DOM depending on the region. This points to interactions between TP, regional hydrology and other, undetermined regional factors that modulate the effect of nutrients on DOM.

Furthermore, we have shown that there is a strong relationship between DOC concentration and DOM molecular composition with d-excess (Figure 2.5). d-excess integrates climatic signatures with hydrologic processes, which underlie lake DOM dynamics at both local and regional scales. Previous studies have reported that DOM concentration varies as a function of precipitation and regional hydrology (Anderson & Stedmon, 2007; Stolpmann et al., 2021), and a few studies in particular have reported a negative relationship between lake DOC concentration and the degree of water evaporation as reflected by water isotopes (Johnston et al., 2020; Kurek et al., 2023; Pugh et al., 2021). A recent study has linked abrupt changes in regional climate and hydrology, also reflected in major shifts in lake water isotopes, with the abrupt onset of browning in Greenland lakes (Saros et al., 2025), further supporting the tight connection that exists between hydrology, aquatic/terrestrial connectivity and DOM dynamics in lakes. Although previous studies had established a negative relationship between DOC concentration in lakes and rivers and water residence time along the entire inland water continuum (Catalán et al., 2016; Kellerman et al., 2014; Köhler et al., 2013), suggesting a continuous removal of DOC along the network, our results at a continental scale suggest the inverse relationship, where DOC concentration tends to increase with the extent of evaporation, and therefore with the residence of the water within the network. Others have reported similar patterns of increasing DOC concentrations as a function of water evaporation for subarctic lakes (Johnston et al., 2020; Kurek et al., 2023).

This apparent discrepancy may originate from the inclusion in our database of lakes that have little hydrologic connectivity and extremely long residence times, such as Prairie lakes (i.e., Hudson Bay basin) and shallow subarctic lakes (i.e., Arctic Ocean basin), which may have extremely high concentrations of mostly photodegraded and aged DOC. It is interesting to note that whereas DOM concentration may increase along a gradient of water evaporation, not all the molecular classes within the DOM behaved in the same manner, and the different classes related to d-excess with different signs (Figure 2.5). In particular, we found that Aromatics and Low.O.unsat classes were selectively removed along a gradient of water evaporation, whereas Aliphatics and High.O.unsat increased as evaporation and the overall residence time of the water increased. This pattern agrees with previous observations for Alaskan lakes (Johnston et al., 2020), and results in major shifts in DOM composition along hydrologic gradients, which explain differences both among lakes and among regions. Network processing is therefore not always

associated with a net decline in DOC, and it depends on the type of lake and network that is assessed, but it appears to be systematically associated to the selective removal of the same molecular moieties, and the addition of the same molecular groups.

We conclude that TP influences DOM via its role in modulating lake primary production and ecosystem metabolism, whereas the hydrologic history of the water, as reflected in d-excess, integrates both the selective DOM degradation and accumulation of DOM that occurs within the aquatic network. Our results suggest that the strength of evaporation (d-excess) together with the source composition of fresh DOM significantly influence the overall composition and characteristics of DOM. This agrees with other published studies that have reported that both source and processing of aquatic DOM vary and are important factors determining DOM composition at regional scales (Johnston et al., 2020; Kellerman et al., 2020; Kurek et al., 2023; Pugh et al., 2021). The composition of DOM that we observe in any given lake therefore reflects a combination of internal lake processes, which shape the transformation the incoming DOM and the production of new DOM, the position of the lake in the network, which shapes the history of the incoming DOM and therefore its composition, and the region where the lake is located, which shapes the composition of the source DOM and the hydrologic patterns.

2.6 Acknowledgments

We thank the lead PI of NSERC-funded LakePulse Project, Yannick Huot, as well as the many technicians, students, and research professionals involved in the collection and generation of data for the network. We would like to thank Mathilde couturier for her contribution to the PARAFAC analysis. We would like to thank the Interuniversity Research Group in Limnology (Groupe de recherche interuniversitaire en limnologie [GRIL]) and their funders, the Fonds de recherche—nature et technologie (FRQNT, Québec). This is part of the research program of the NSERC / HydroQuébec Research Chair in Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBas) led by PdG and which also supported AS. We would like to thank Alice Parkes for coordinating all the laboratory work. We also would like to thank the Water Quality Centre, Trent University, Peterborough, Canada for the use of their instrumentation in DOM composition analysis.

2.7 Data availability

The data supporting the findings of this study are available in the Borealis – Canadian Dataverse Repository [Shahabinia et al., 2025]. The dataset can be accessed at https://doi.org/10.5683/SP3/LZAUGS

2.8 Supplementary Information

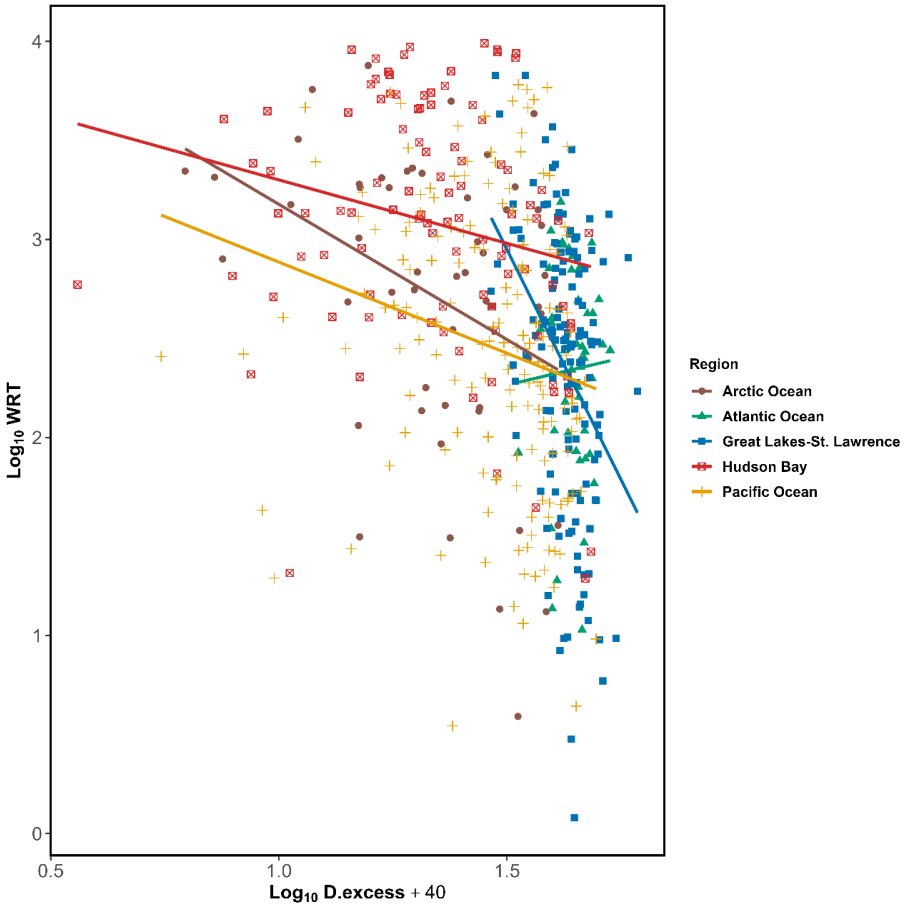

Figure S2.1: Relationship between estimated water residence time (WRT) and measured d-excess across all lakes and regions.

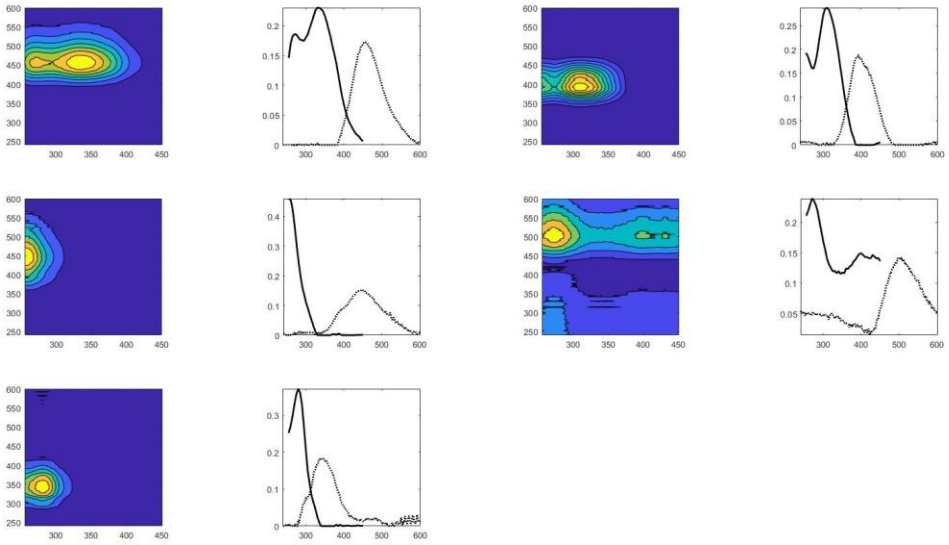

Figure S2.2: Five independent fluorescent components identified using PARAFAC analysis.

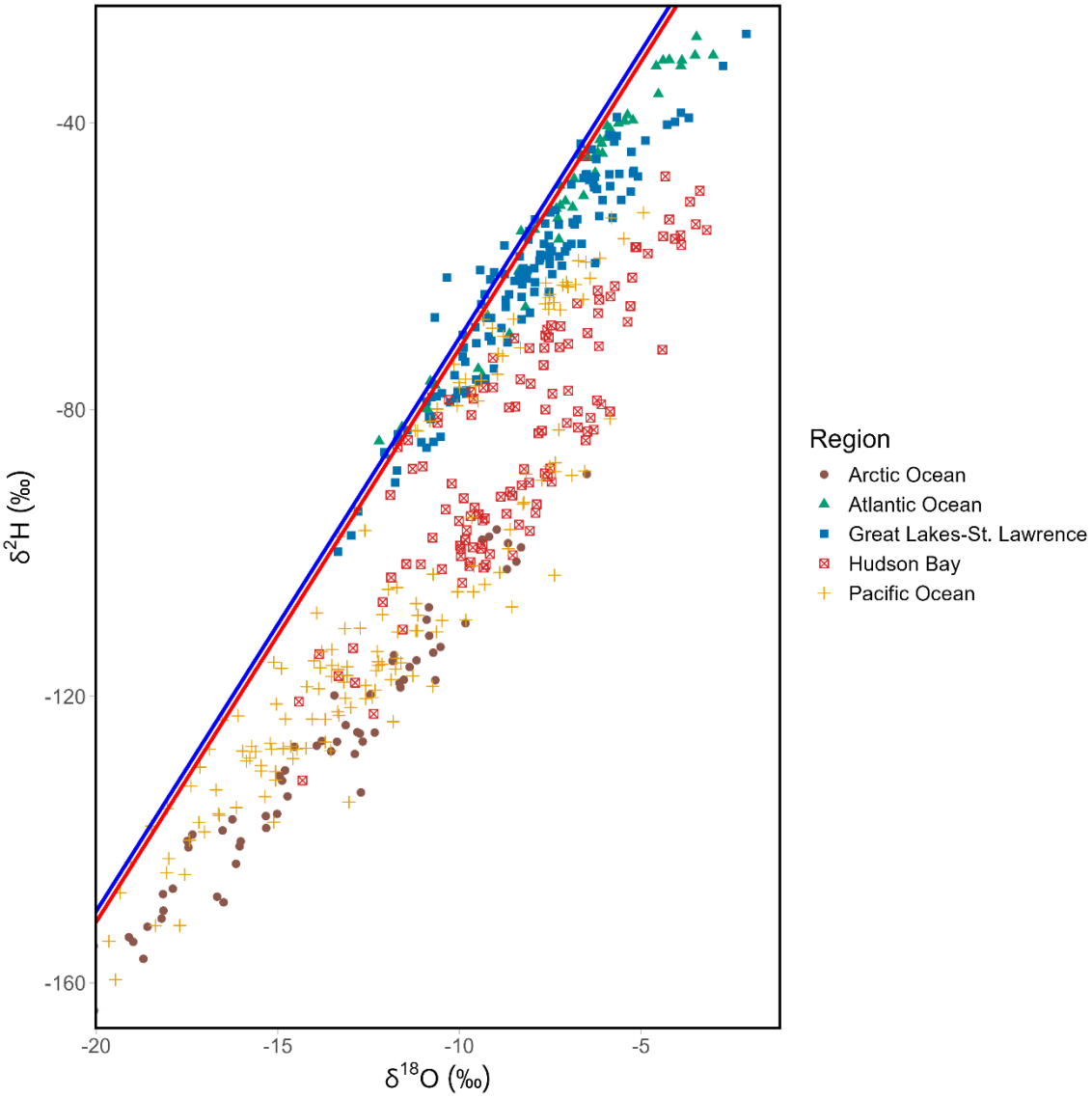

Figure S2.3: Relationship between $\delta^{18}O$ versus $\delta^{2}H$ in lakes across Canada. Global Meteoric Water Line (solid blue line), the Canadian Meteoric Water Line (CMWL) (solid red line).

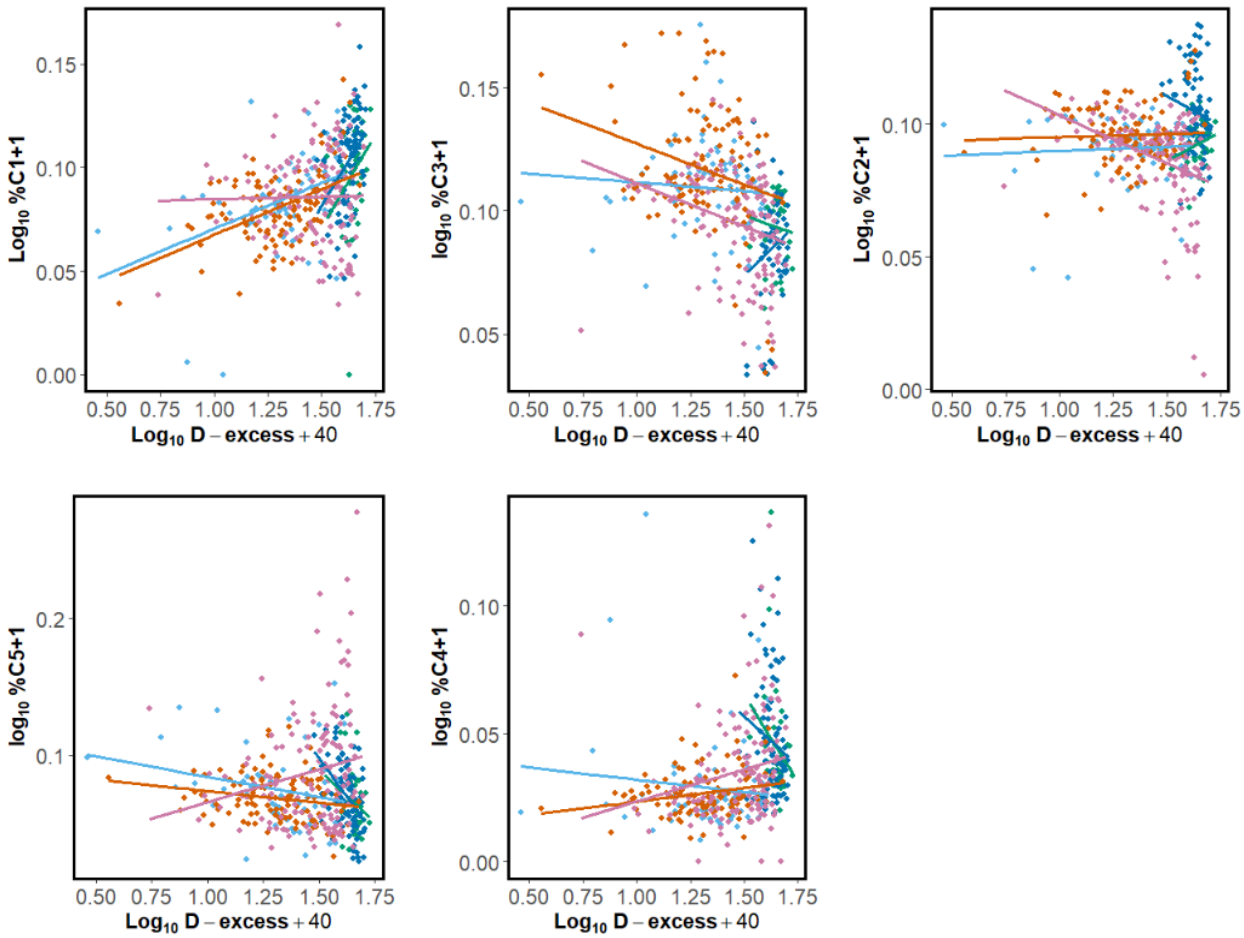

Figure S2.4: Relationships between PARAFAC components (%C1, %C2, %C3, %C4 and %C5) and d-excess across different regions.

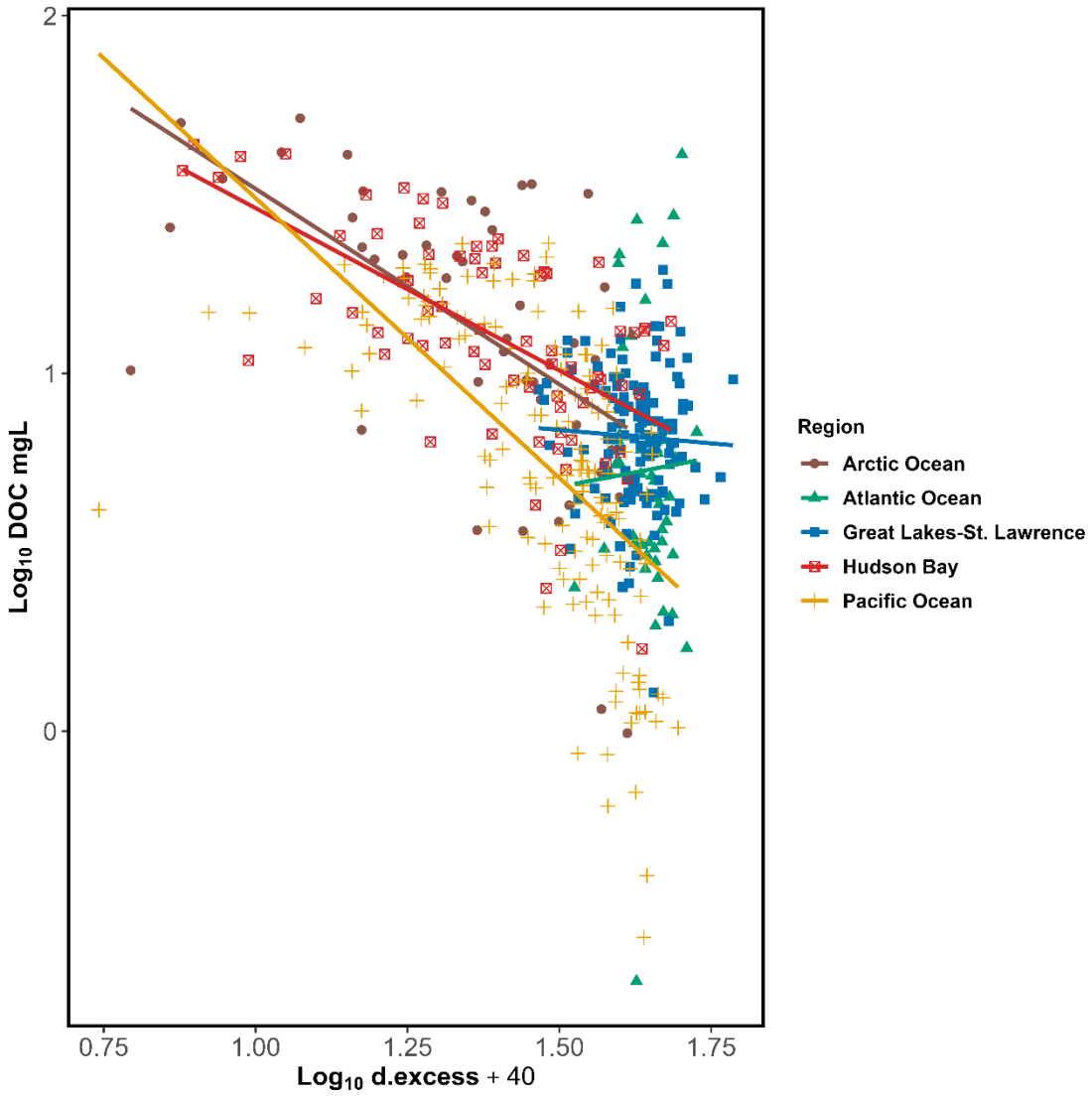

Figure S2.5: Relationships between DOC concentration and d-excess across different regions.

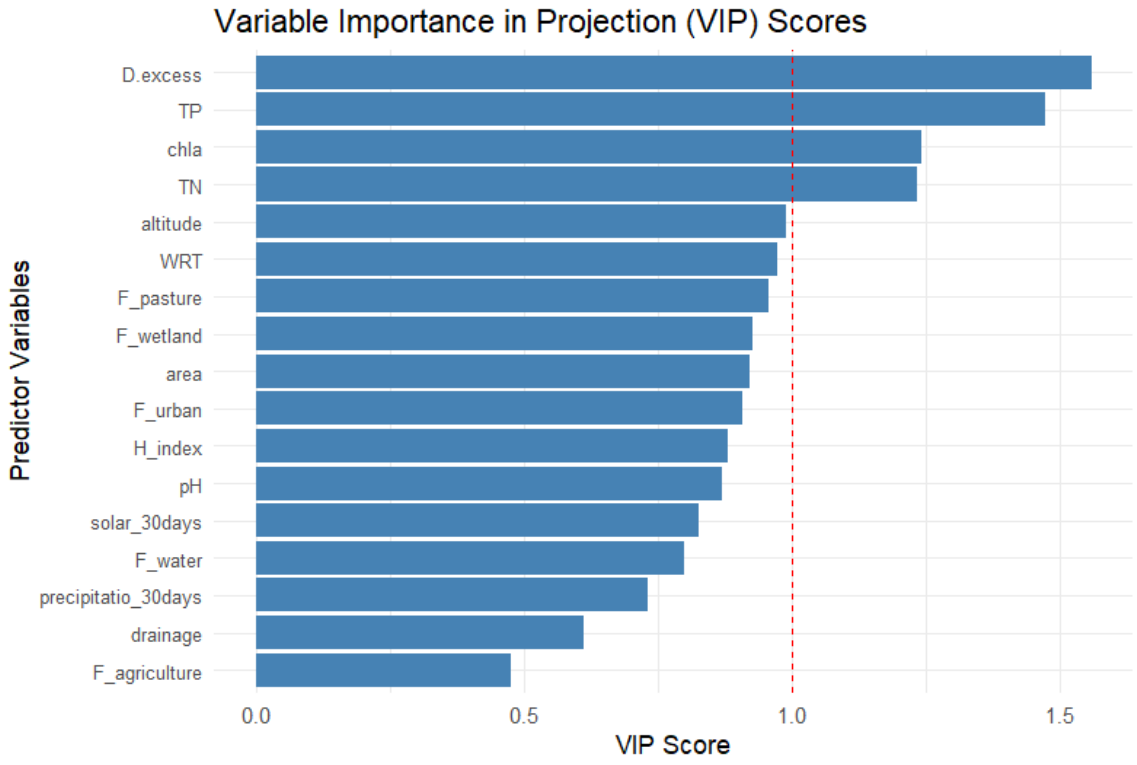

Figure S2.6: Variables importance in projection (VIP) figure calculated based on PLS analysis.

Table S2.1: Detailed information regarding environmental and watershed characteristics used in this study.

Parameter	Abbreviation	Unit	Median	Mean	std.dev	Range
Lake area	area	Km^2	0.6	9.9	16.3	0.007 - 266
Catchment area	catchment	Km^2	17.9	288	1775	0.2 - 37459
Maximum depth	Zmax	m	6.6	5.2	10.05	1 – 64.1
Drainage area	drainage	n.a	14.59	62.35	10.4	1.54 – 2429.6
Chlorophyll a	chla	μgL^{-1}	2.9	6.8	10.7	0.04 - 72
Total phosphorus	TN	μgL^{-1}	1.8	29.2	23.9	2 – 4.5
Total nitrogen	TN	mgL^{-1}	0.2	0.58	0.8	0.01 - 11.9
Dissolved organic carbon	DOC	mgL^{-1}	8.2	13	12	0.20 - 220
Colored dissolved organic matter	CDOM	m^{-1}	1.12	1.65	1.7	0.006 – 9.7
Water residence time	WRT	day	376.8	1098.04	1579	0.1 - 9776
deuterium excess	D.excess	‰	-5.42	-7.42	13.3	-37.1 – 21.1
Altitude	altitude	m	359.5	440.3	302	2 - 1533
30 days period solar radiation	solar_30days	$J\ m^{-2}\ day^{-1}$	561175	569049	6532	4052577 - 7724438
30 days period precipitation	precipitatio_30days	m	0.97	0.99	0.54	0.04 – 2.58
Fraction of agriculture	F_agriculture	%	0.001	0	0.18	0 – 0.022
Fraction of urban	F_urban	%	0.018	0.068	0.13	0 - 0.88
Fraction of wetland	F_wetland	%	0.0014	0.019	0.04	0 - 0.49
Fraction of water	F_water	%	0.11	0.13	0.09	0.0008 - 0.66
Fraction of pasture	F_pasture	%	0.0011	0.066	0.10	0 - 0.59
Human impact index	H_index	n.a	0.11	0.22	0.3	0 – 0.97

Table S2.2: Result of models for DOC-CDOM relationship across continental basins.

Model	Region	Intercept	Slope	P-value
DOC-CDOM	Arctic Ocean	-0.95	0.91	0.001***
	Atlantic Ocean	-0.49	0.72	0.001***
	Hudson Bay	-0.54	0.57	0.74
	Great Lakes St. Lawrence	-1.001	1.24	0.003**
	Pacific Ocean	-1.11	1.15	0.0018*

Significant : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table S2.3: Results of multiple regression models of DOC, molecular and optical properties based on d-excess, TP, and basin as a qualitative variable.

Model	Variable	Coefficient	St. Error	P-value
DOC= (d-excess) ~ (TP)*Region	Log10 d-excess	-0.76	0.24	0.001***
	Log10 TP	0.78	0.14	0.001***
	Arctic Ocean	1.14	0.24	0.001***
	Atlantic Ocean	0.62	0.25	0.01*
	Hudson Bay	0.49	0.22	0.03*
	Great Lakes St. Lawrence	0.69	0.21	0.001**
	Pacific Ocean	0.02	0.2	0.9
Aromatic= (d-excess) ~(TP)*Region	Log10 d-excess	0.71	0.06	0.001***
	Log10 TP	0.29	0.09	0.01**
	Arctic Ocean	0.57	0.22	0.08
	Atlantic Ocean	0.28	0.14	0.09.
	Hudson Bay	0.13	0.14	0.34
	Great Lake St. Lawrence	0.28	0.15	0.04*
	Pacific Ocean	0.22	0.12	0.08.
Aliphatic= (d-excess) ~(TP)*Region	Log10 d-excess	-0.31	0.04	0.001***
	Log10 TP	-0.19	0.07	0.01*
	Arctic Ocean	1.44	0.13	0.001***
	Atlantic Ocean	-0.11	0.13	0.38
	Hudson Bay	-1.14	0.12	0.23
	Great Lakes St. Lawrence	-1.11	0.11	0.34
	Pacific Ocean	-0.04	0.1	0.71
HighO.unsat= (d-excess) ~(TP)*Region	Log10 d-excess	-0.12	0.01	0.001***
	Log10 TP	0.03	0.01	0.07.
	Arctic Ocean	1.84	0.03	0.001***
	Atlantic Ocean	0.05	0.03	0.09.
	Hudson Bay	0.02	0.03	0.37
	Great Lakes St. Lawrence	0.03	0.02	0.27
	Pacific Ocean	0.02	0.02	0.37
LowO.unsat= (d-excess) ~(TP)*Region	Log10 d-excess	0.08	0.01	0.001***
	Log10 TP	-0.07	0.02	0.002**
	Arctic Ocean	1.50	0.04	0.001***
	Atlantic Ocean	-0.1	0.04	0.016*
	Hudson Bay	-0.03	0.03	0.39
	Great Lakes St. Lawrence	-0.11	0.03	0.002**
	Pacific Ocean	-0.07	0.03	0.02*
%C1= (d-excess) ~(TP)*Region	Log10 d-excess	0.36	0.04	0.001***
	Log10 TP	0.09	0.03	0.01*
	Arctic Ocean	-1.28	0.08	0.01**
	Atlantic Ocean	0.03	0.05	0.59
	Hudson Bay	0.09	0.06	0.15
	Great Lakes St. Lawrence	-0.06	0.07	0.37
	Pacific Ocean	-0.13	0.04	0.05**
%C2= (d-excess) ~(TP)*Region	Log10 d-excess	0.01	0.02	0.53
	Log10 TP	0.05	0.02	0.03*
	Arctic Ocean	-0.72	0.05	0.001***
	Atlantic Ocean	0.07	0.04	0.07.
	Hudson Bay	0.1	0.04	0.01*
	Great Lakes St. Lawrence	0.08	0.05	0.11
	Pacific Ocean	-0.08	0.03	0.01*
%C3= (d-excess) ~(TP)*Region	Log10 d-excess	0.15	0.07	0.03*
	Log10 TP	0.23	0.07	0.0007***
	Arctic Ocean	-1.04	0.15	0.001***
	Atlantic Ocean	0.19	0.11	0.07.
	Hudson Bay	-0.06	0.12	0.58

	Great Lakes St. Lawrence	-0.29	0.13	0.02*
	Pacific Ocean	-0.03	0.09	0.69
%C4= (d-excess) ~(TP)*Region	Log10 d-excess	-0.015	0.06	0.01*
	Log10 TP	-0.3	0.05	0.001***
	Arctic Ocean	-0.66	0.12	0.001***
	Atlantic Ocean	-0.09	0.09	0.32
	Hudson Bay	-0.02	0.1	0.82
	Great Lakes St. Lawrence	0.08	0.11	0.44
	Pacific Ocean	0.03	0.07	0.66
%C5= (d-excess) ~(TP)*Region	Log10 d-excess	-0.47	0.05	0.001***
	Log10 TP	-0.34	0.05	0.001***
	Arctic Ocean	0.24	0.1	0.01*
	Atlantic Ocean	-0.3	0.08	0.001***
	Hudson Bay	-0.12	0.08	0.16
	Great Lakes St. Lawrence	-0.17	0.09	0.06.
	Pacific Ocean	0.04	0.06	0.48

Significant : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table S2.4: Result of models for relationship between different molecular classes across (Aromatic, Aliphatic, High.O.unsat, and Low.O.unsat) with d-excess across different regions.

Model	Region	Intercept	Slope	P-value
Aromatic-d.excess	Arctic Ocean	0.17	0.44	0.27
	Atlantic Ocean	0.70	0.17	0.59
	Hudson Bay	-0.29	0.76	0.02*
	Great Lakes St. Lawrence	-1.04	1.24	0.006**
	Pacific Ocean	-0.20	0.69	0.065.
Aliphatic-d.excess	Arctic Ocean	1.10	-0.23	0.001***
	Atlantic Ocean	2.30	-0.94	0.13
	Hudson Bay	1.03	-0.19	0.63
	Great Lake St. Lawrence	2.60	-1.12	0.001***
	Pacific Ocean	0.82	-0.05	0.048*
High.O.unsat-d.excess	Arctic Ocean	1.83	-0.08	0.001***
	Atlantic Ocean	1.41	0.16	0.044*
	Hudson Bay	1.95	-0.14	0.001***
	Great Lakes St. Lawrence	1.76	-0.04	0.047
	Pacific Ocean	1.95	-0.16	0.001**
Low.O.unsat-d.excess	Arctic Ocean	1.38	0.08	0.001***
	Atlantic Ocean	1.22	0.17	0.569
	Hudson Bay	1.25	0.14	0.004**
	Great Lakes St. Lawrence	1.67	-0.1	0.017*
	Pacific Ocean	1.30	0.14	0.079.

Significant : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CHAPITRE 3

Dissolved organic matter (DOM) pools modulate the relationship between production and respiration in lakes

Amir Reza Shahabinia[1], Jean-François Lapierre[2,3], Paul A. del Girogio[1]

[1] Department of Biological Sciences, University of Quebec at Montreal, Montreal, Quebec, Canada

[2] Interuniversity Research Group in Limnology (Groupe de Recherche Interuniversitaire en Limnologie; GRIL), Montreal, Quebec, Canada

[3] Department of Biological Sciences, University of Montreal, Montreal, Quebec, Canada

Keywords: Lake metabolism, Dissolved organic matter, Structural equation model, path analysis,

N.B. References cited in this chapter are presented at the end of the thesis.

3.1 Abstract

Lake metabolism and dissolved organic matter (DOM) concentration and composition known to play a crucial role in shaping ecosystem processes and C dynamics in lakes. Although the dynamics of both Lake metabolism and dissolved organic matter (DOM) concentration and composition are relatively well-studied, less is known about the links between metabolism and DOM composition. Here, we present the results of a large-scale study, where we explore the potential bidirectional links between DOM and metabolism in 548 lakes across Canada. We measured summer metabolic rates in these lakes using an oxygen isotopic approach. We also studied the quantity and quality of DOM using absorption spectroscopy, parallel factor analysis modeling, and ultra-high resolution mass spectroscopy. The sampled lakes were distributed across Canada, covering a wide range of in-lake, watershed, and climatic features. We used structural equation modelling to test potential relationships between DOM and metabolism and of their responses to environmental forcing. GPP and R varied as a function of TP and DOC levels. DOC concentration was directly related to water history as evidenced by d-excess and indirectly to watershed and climate properties. There was a tight coupling between GPP and R along the productivity gradient. GPP appeared to be positively related to certain DOM components such as C5 and Aliphatic and negatively to C1, Aromatic and High.O.unsat. In contrast High.O.unsat class was positively related to R. Our findings suggest that not all the pools within the bulk lake DOM are influenced by lake metabolism, and conversely, not all the DOM pools have the capacity to influence lake metabolism.

3.2 Introduction

Lake metabolism plays a major role in ecosystem functioning, and quantifying gross primary production (GPP), respiration (R) and net ecosystem production (NEP = GPP-R) and identifying their main drivers are keys to our capacity to understand lake functioning and the role of lakes in the landscape carbon (C) balance. Among the many variables influencing lake metabolism, dissolved organic carbon (DOC) has been shown to play a central role (del Giorgio et al. 1999; Solomon et al. 2013; Bogard et al. 2020; Holgerson et al. 2022). Exploring the relationship between lake metabolism and dissolved organic matter (DOM) has been challenging because whereas DOM influences lake metabolism, various aspects of lake metabolism also shape lake DOM, thus creating a complex, bidirectional connection that has been difficult to unravel (Smith et al. 2018).

In order to disentangle and understand these complex relationships between DOM and lake metabolism, it is necessary to go beyond DOC concentration. The DOM is composed of an extremely complex mixture of molecules that vary greatly across inland waters. The chemical and optical composition of DOM in lakes is partly the result of adding, removing and transforming of moieties by processes such as GPP and R within the lake (Stadler et al. 2020). This implies that lake metabolism is likely associated to the production of certain DOM pools and the removal of others, and that these DOM dynamics will depend on the intensity of metabolic processes, and thus on system productivity and trophic status (Smith et al. 2018). Yet much of the DOM that circulates through lakes, particularly in northern landscapes, is of terrestrial origin, and the amount and chemical composition of this allochthonous DOM influences lake function in multiple ways. Colored DOM can suppress GPP by altering underwater light regimes and reducing thermocline depth (del Giorgio et al. 1994; Karlsson et al. 2009), yet terrestrially-derived DOM can enhance GPP by delivering nutrients (Tanentzap et al. 2017), and certain allochthonous DOM pools may enhance R by providing C sources to bacteria and other heterotrophic organisms. Whether it is through suppressing or enhancing GPP and R, allochthonous DOM can potentially influence whole-lake NEP (Ask et al. 2012; Zwart et al. 2016; Oleksy et al. 2024). Yet other terrestrially derived DOM pools that are uncolored, unliked to nutrient or biologically unreactive may neither influence metabolism, nor be influenced by it. Overall, it is clear that not all the pools within the bulk lake DOM are influenced by lake metabolism, and conversely, not all the DOM pools have the capacity to influence lake metabolism.

Although there have a been several efforts to address the DOM/metabolism relationship (e.g., Osburn et al., 2011, Ayala-Borda et al. 2024), our understanding of the links between DOM and lake metabolism

remains limited. To disentangle this complex, bidirectional relationship between DOM and metabolism we must simultaneously consider a range of scenarios that reflect the diversity of lake types, catchment characteristics, and climatic conditions, all of which influence DOM concentration and chemical composition and also influence lake metabolism. In particular, it would be necessary to assess the DOM/ metabolism relationship along a gradient of varying levels of ecosystem metabolic rates, from unproductive, oligotrophic to highly productive eutrophic lakes to assess potential shifts in the production of certain DOM poos. Likewise, it would be important to assess this relationship across lakes that vary in the net metabolic balance between GPP and R, from net autotrophic (GPP>R) to net heterotrophic lakes, since the interactions with DOM may shift as a function of metabolic balance. Finally, it would be necessary to consider lakes differing in DOM origin and therefore composition, from those dominated by allochthonous versus autochthonous sources. Generating these scenarios and contrasts would help disentangle the relationship between DOM and lake metabolism and identify what pools are involved in these complex interactions.

Here we present a continental-scale assessment of lake metabolism and DOM concentration and composition across Canadian lakes, where we seek to determine the links between lake metabolism components and DOM composition and concentration. The data collected from 548 lakes across 11 ecoregions covering extremely wide climatic, geographic and human impact gradients provided a unique opportunity to test a diverse set of scenarios with lakes spanning a wide range of ecosystem metabolic rates and metabolic balance, and, in also in DOM origin, concentration and composition. We applied a combination of advanced analytical tools to assess DOM concentrations and composition and lake metabolism. We assessed the optical properties of DOM based on excitation emission spectra, and the molecular properties of DOM using Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). We assessed summertime GPP, R, and their balance, NEP in the same set of lakes using the oxygen isotopic ($\delta^{18}O_2$) free-water approach. Here we address four key questions: 1) How does DOC influence GPP and R across Canadian lakes? 2) What are the climatic and environmental factors mediating these relationships? 3) What are the pools of DOM that influence GPP and R? 4) What pools of DOM are influenced by GPP and R.

Table 3.1: climatic, land-use and in-lake characteristics of the lakes studied in this chapter.

Parameter	Abbreviation	Unit	Median	Mean	St.dev	Range
Maximum Depth	max_depth	m	4.6	8.5	9.86	1 – 54.7
Lake area	area	Km^2	0.6	5.2	16.3	0.008 – 91.5
Chlorophyll a	chla	μgL^{-1}	3.4	7.9	12.0	0.04 - 72
Total phosphorus	TP	μgL^{-1}	20.5	30.1	162.8	3 - 150
Total nitrogen	TN	mgL^{-1}	0.3	0.7	0.9	0.01 – 4.39
Dissolved organic carbon	DOC	mgL^{-1}	12.5	18.3	12	1.02 - 220
Dissolved inorganic carbon	DIC	mgL^{-1}	23.9	41.3	63.04	0.02 - 432
Water temperature	water_ temperature	C	20.6	20.5	2.3	13 - 25
Water column light	Water_light	umol photo m^{-2} s^{-1}	238.3	250.1	0.13	31.05 – 777.6
Water residence time	WRT	day	410.4	1214.7	1867.1	1 - 9088.4
Deuterium excess	d-excess	‰	-10.04	-9.93	0.71	-37.26 – 13.2
pH	pH	n.a	8.3	7.6	2.3	1 - 10
Catchment	catchment	Km^2	12.8	444.4	2653.1	0.2 - 37459.8
Fraction of urban	F_urban	%	0.02	0.06	0.14	0 - 0.92
Fraction of forest	F-forest	%	0.001	0.02	0.59	0 - 0.5
Fraction of water	F_water	%	0.1	0.13	0.17	0.0006 - 0.66
Fraction of pasture	F_pasture	%	0.005	0.054	0.08	0 - 0.4
Fraction of agriculture	F_agriculture	%	0.001	0.13	0.22	0 – 0.86
Fraction of natural land	F_natural	%	0.65	0.59	0.27	0 – 0.98
Human impact index	H_index	n.a	0.14	0.27	0.28	0 – 0.97
Altitude	altitude	m	495	485.6	298.5	8 - 1513
Latitude	Latitude	°	n.a	n.a	n.a	43.57 – 62.55
30 days period solar radiation	solar_30days	J m^{-2} day^{-1}	56464	57093	6505	42094 - 72171
30 days period precipitation	precipitation_30days	m	0.97	0.99	0.52	0.04 - 2.3
Temperature	temperature	C degree	118.6	18.78	4.17	8 – 29.4
Wind	wind	m s^{-1}	2.32	2.44	0.93	0.6 – 5.75
Humidity	humidity	%	70.90	76.69	20.6	0 - 100
Pressure	pressure	hPa	978.9	970.76	34.6	853.1 - 1024.1

3.3 Material and Methods

3.3.1 Sampling and environmental characterization

We sampled 548 lakes between 2017 and 2019, from late June to early September across Canada as part of the Canadian LakePulse Network, funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) (Huot et al. 2019) (Figure 3.1a). Lakes were visited once to sample a range of limnological parameters within a single day, and one lake was sampled per day during the field campaigns. Epilimnetic samples were obtained at the deepest point of each lake using an acid-washed 2-meter tube sampler and kept in dark and cold for further analysis. Water temperature, dissolved oxygen, pH, salinity, conductivity, and chlorophyll-a vertical profiles were collected throughout the epilimnion or the whole water column in the case of shallow lakes using an RBR Maestro multi-parameter water quality meter (RBR Ltd., Ottawa, Canada). Hourly wind data during the sampling period (7 days prior to sampling), precipitation, and solar

radiation (30 days prior to sampling) were obtained from the nearest meteorological station in each region. A summary of the limnological and catchment properties used in this study is presented in Table 3.1.

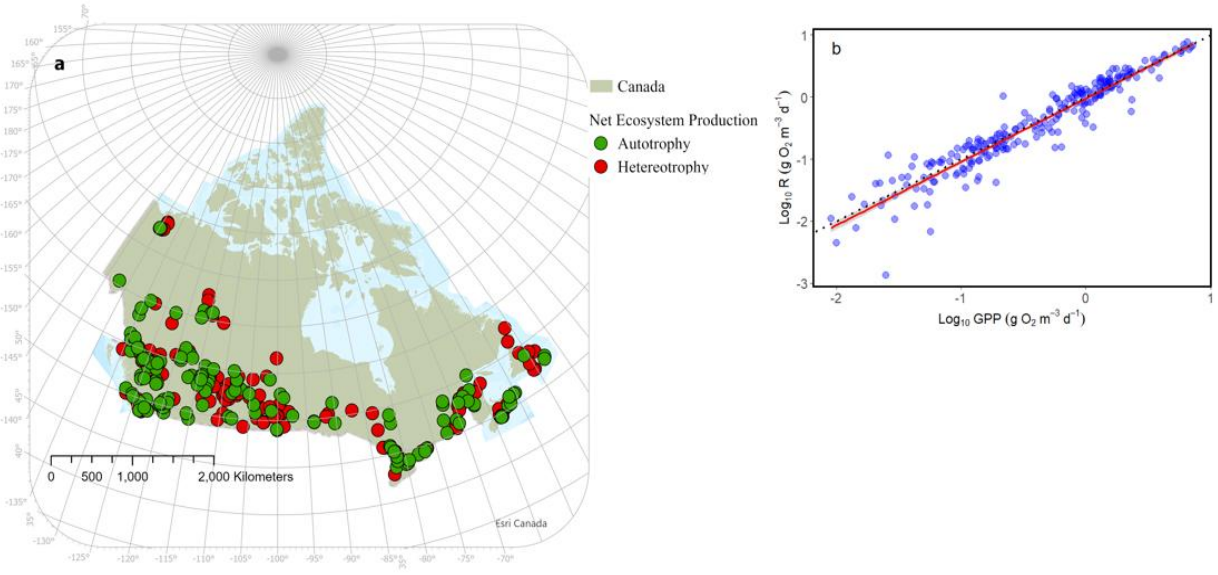

Figure 3.1: a) Map of sampled lakes by the NSERC Canadian LakePulse Network. b) Relationship between GPP and R. dotted line is 1:1 line and the red line is the regression line.

Triplicate samples for $\delta^{18}O_2$ analysis were collected from the surface waters at the deepest point in each lake and preserved in 12 mL vials with 120 µl of saturated $ZnCl_2$ solution. For isotopes of hydrogen (δ^2H) and oxygen $\delta^{18}O_2$ in water, samples (H_2O) were taken from the same point in the surface water of each lake, stored in 30 mL HDPE bottles without any air bubbles, and subsequently kept at 4℃ in the dark until further analyses. Deuterium excess (d-excess) was calculated as follows: d-excess = $\delta^2H - 8 \times \delta^{18}O$ (Dansgaard, 1964). d-excess represents a departure from the local meteoric signal and as such can be used as an index of the degree of evaporation in a given water sample. To estimate summertime mixed-layer metabolic rates in lakes, we used an oxygen isotope ($\delta^{18}O_2$) approach, which combines ambient O_2 concentration and its isotopic signature, as detailed in (Bogard et al. 2017, Supplementary 3.8.2: Oxygen isotope method). Water samples for DOM concentration and its optical and molecular properties were filtered using a 0.45 µm syringe filter in the mobile lab on the same day of sampling. In the laboratory, filtered water samples were used to measure absorption (230 to 700 nm) using a Biochrom Ultrospec 3100 Pro spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA), to derive CDOM as absorbance

at 440nm. To build three-dimensional excitation and emission matrices (EEMs), fluorescence intensity was measured using a spectrofluorometer (Shimadzu RF5301 PC, Shimadzu, Kyoto, Japan) across excitation wavelengths of 230–450 nm (5 nm increments) and emission wavelengths of 240–600 nm (2 nm increments). Parallel factor analysis (PARAFAC) was used to characterize the DOM fluorescence signal using MATLAB software (MATLAB 7.7.0, The MathWorks, Natick, USA, 2021) and drEEM toolboxes (Stedmon and Bro, 2008; Murphy et al., 2013). The parameters obtained from the PARAFAC model were used to calculate an approximation of the abundance of each component, expressed as Fmax (in R.U.), which corresponds to the maximum fluorescence intensity for a particular sample. Additionally, to investigate the intensity of each PARAFAC component (Ci), the relative fluorescence intensity of each component as a percentage (% Ci) was calculated (% Ci = Ci/ΣCi * 100%). The PARAFAC components were matched to previously reported components using the OpenFluor database (Murphy et al., 2014). PARAFAC analysis of EEMs identified five optical DOM components: Component 1 (C1) was a humic-like component (Kothawala et al., 2012); component 2 (C2) was characterized as substances altered by microbial reprocessing (Lapierre & 169 del Giorgio, 2014); component 3 (C3) was characterized as terrestrial humic-like (Kothawala et al., 2014); component 4 (C4) was considered fulvic acid-like (Søndergaard et al., 2003); component 5 (C5) was characterized as protein-like (Lapierre & del Giorgio, 2014). More details on the PARAFAC components can be found in supplementary Figure S3.2. The molecular composition of DOM was analyzed in filtered water samples by Fourier transform-ion cyclotron resonance mass spectrometry (FT-ICR MS). Solid-phase extraction was carried out using a modified styrene divinyl benzene polymer-type sorbent (100 mg Bond Elut PPL cartridges, Agilent Technologies) (Dittmar et al., 2008) and vials were kept in a dark freezer at -20°C until further analysis. The DOM solid-phase extracts were analyzed using a 7T Bruker SolariX XR (FT-ICR MS) at the Water Quality Centre, Trent University, Peterborough, Canada. Details of data processing and cleaning are provided in Supplementary 3.8.1: Ft-ICR MS analysis and in chapter 2). We used four broad chemical compound classes, based on O/C, H/C, and modified aromaticity index (AImod), which quantifies the degree of aromaticity of an organic compound (Koch and Dittmar, 2006). The classes were "aliphatic" (H/C ≥ 1.5), "Low.O.unsaturated" (H/C < 1.5, AImod < 0.5, O/C < 0.5), "High.O.unsaturated" (H/C < 1.5, AImod < 0.5, O/C $\geqslant$ 0.5), and "aromatics" (0.5 < AImod < 0.67). From a total of 548 lakes, complete datasets were retained for 521 lakes in ecosystem metabolism analyses, 548 lakes in FT-ICR MS measurements, and 400 lakes in PARAFAC modeling.

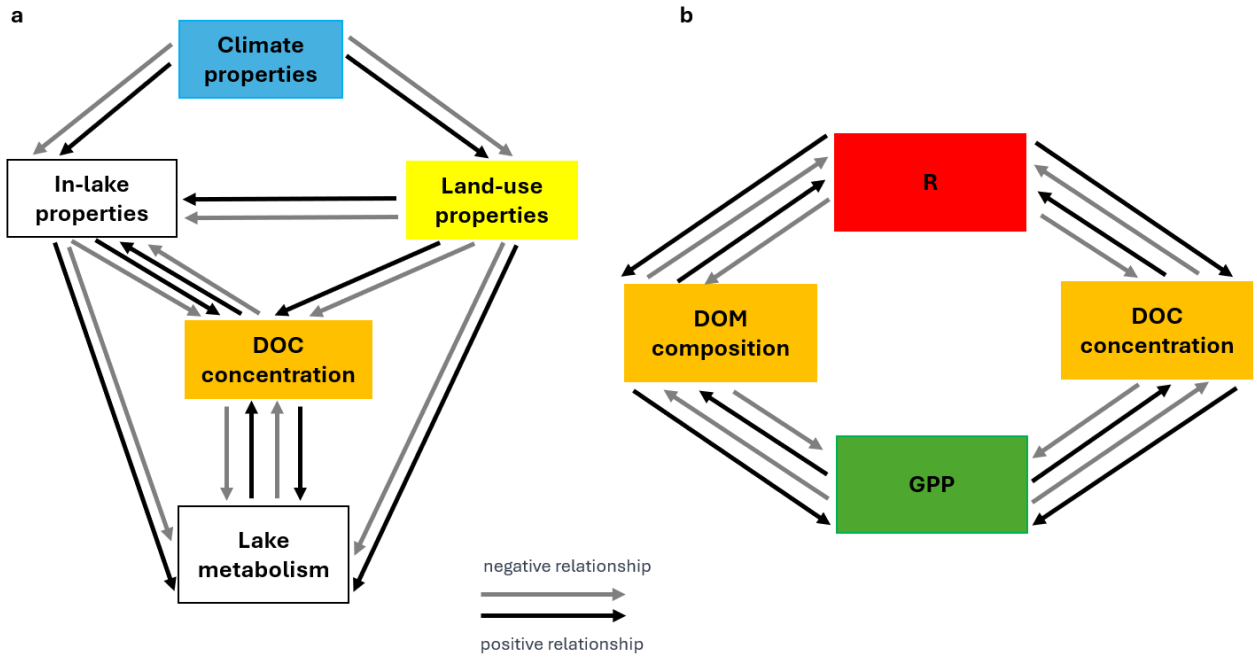

Figure 3.2: Conceptual figure Structural equation model (SEM) used in the path analysis. a) Relationship between climate properties, land-use and in-lake properties which might have positive or negative b) Arrows indicate the direction of causality.

We conducted checks on the normality of metabolic, and other environmental data. When necessary, we transformed the data to log and for fractional watershed variables, a square root approach was applied. Statistical tests were conducted using R v. 4.1.0 (R Core Team 2021). To investigate how environmental variables relate to each other and more importantly to reduce and summarize collinear variables, we divided environmental variables into three main categories: in-lake, watershed, and climate variables and we performed principal components analysis (PCA) with the *ggfortify* package's *precomp()* function. Prior to PCA, data was scaled and centered. We performed non-metric multidimensional scaling (NMDS) with the *vegan* package's *metaMDS()* function to visualize and analyze relationships between different DOM components. We used a structural equation modeling (SEM) framework to test potential relationships between DOM and lake metabolism, and the potential responses to both to environmental factors. SEM allowed us to test hypothesized directional patterns (MacCallum and Austin 2000), and we used the basic framework presented in Figure 3.2a to guide our path analysis. In this framework, regional climatic conditions may influence both in-lake and catchment properties, which in turn may influence both lake

metabolism and DOM concentration and composition. The DOM concentration and composition may influence one or more aspects of lake metabolism and vice versa, positively or negatively (figure 3.2b). To find the most parsimonious SEM that could explain our data, we used an iterative fitting procedure, at each step we either added a link that improved overall model fit, or removed a link that reduced overall model fit. The fit of each SEM to data was assessed using the comparative fit index (CFI), the Root Mean Square Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR). SEM analyses were carried out with the *Lavaan* package (Rosseel 2012) *sem()* function.

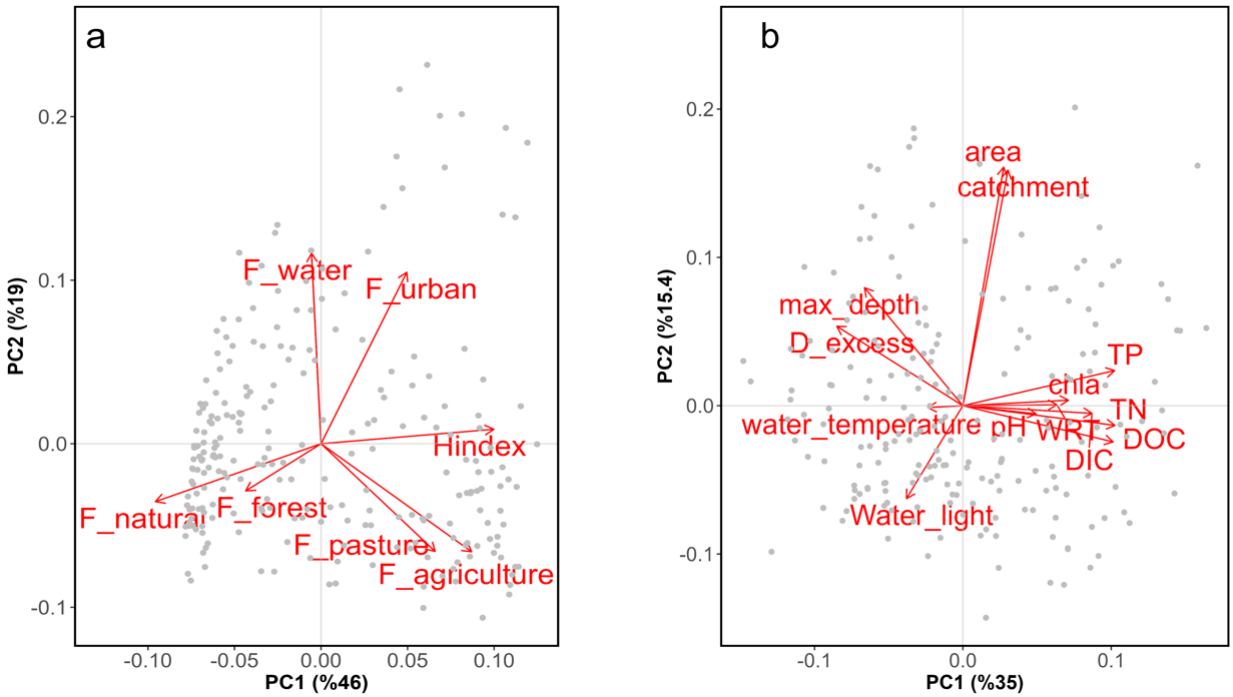

Figure 3.3: Principal component analysis (PCA) on land use and in-lake properties for the sampled lakes. Abbreviations: F_water – fraction of water, Hindex – human impact index, F_natural – fraction of natural land, F_urban – fraction of urban, F_forest – fraction of forest, F_pasture – fraction of pasture, F_agriculture – fraction of agriculture. DOC – dissolved organic carbon, TP – total phosphorus, TN – total nitrogen, chla - chlorophyll a, area – lake area, WRT – water residence time, D-excess – deuterium excess, catchment – catchment, water_ temperature – water temperature, pH – pH, Water_light – light in the water column, max_ depth – maximum depth

3.4 Results

The studied lakes and their watersheds covered a wide range of land use, climate, and lake properties, as shown in Table 3.1. To explore the spatial patterns in environmental variables, we conducted a PCA

analysis with the variables separately categorized into land use, in-lake (Figure 3.3), and climate properties (Figure S3.2). Figure 3.3a shows the PCA distribution of lakes based on land use and human impact variables, where the first two axes together explained 61% of the variability, with fraction of agriculture, pasture and natural lands weighing strongly on PCA1, and fraction of urban and water weighing on PCA2. Lakes across Canada are clearly separated based on land use, and in particular lakes with high human impact and a large fraction of agriculture and pasture in their watersheds are clearly separated from lakes with a large fraction of natural landscape and forest. PCA that corresponds to in-lake features (Figure 3.3b) distinguished with nutrients, DOC and d-excess weighing strongly on PCA1 separating and catchment and lake area, light weighing on PCA2. Across all lakes, summer GPP and R ranged 4 orders of magnitude, from 0.009 to 7.2 g O_2 m^{-3} day^{-1}, and 0.001 to 7.6 g O_2 m^{-3} day^{-1}, for GPP and R, respectively, and there was a strong positive relationship between both (log slope of 0.99; Figure 3.1b).

To assess how patterns of the major chemical classes and optical components of DOM varied across lakes in Canada, we conducted a non-metric multidimensional scaling (NMDS) (Figure 3.4). The NMDS shows a strong coherence between major chemical classes and the main optical components derived from PARAFAC analyses, with each major chemical class strongly aligning to one of the optical components. NMDS axis 1 is dominated by variations in the proportion of Aromatics (and C1), and opposingly the proportion of Aliphatic (and C5 optical component), whereas NMDS axis 2 is dominated by variations in High.O.unsat versus Low.O.unsat molecular classes, and in C3 versus C4 optical components. Overall, as lakes become impoverished in Aromatic (and C1), they tend to become enriched in Aliphatic (and C5) and high.O.unsat (C3).

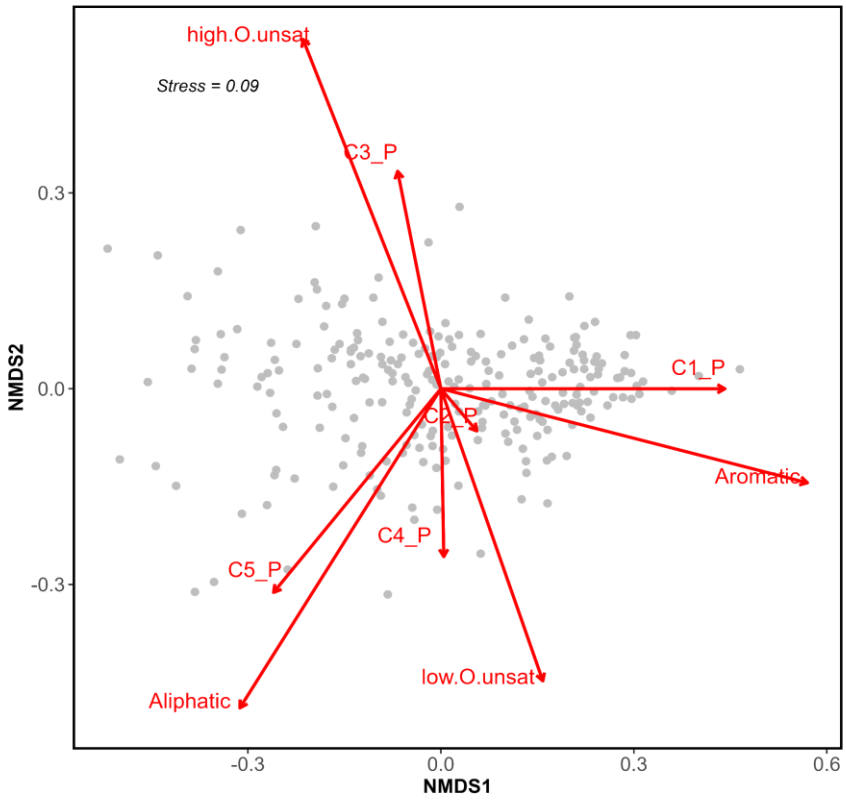

Figure 3.4. Non-metric Multidimensional Scaling (NMDS) for molecular and optical components of DOM for lakes across all sampled lakes. Abbreviation: Aromatic – Aromatic, Aliphatic – Aliphatic, C1_P – %C1, C2_P – %C2, C3_P – %C3, C4_P – %C4, C5_P – %C5, high.O.unsat – High.oxygen.unsaturated, low.O.unsat – Low.oxygen.unsaturated.

We conducted a SEM path analysis following the basic structure presented in Figure 3.2, relating GPP and R to each other and to the lake properties, (d-excess, TP and DOC), which are in turn related to both land-use and climate properties (Figure 3.5). We used the first-axis scores from the PCA of land-use metrics (Figure 3.3a) to integrate these colinear variables in our path analysis. The first two components of the PCA captured 65% of the variance in the dataset (axis 1 captured 46%), which included fraction of water, fraction of natural land, fraction of pasture, fraction of agriculture, fraction of forest, fraction of urban and human impact index. The resulting SEM fits very well (CFI = 0.972, RMSEA = 0.087, SRMR = 0.055) in index values proposed by Fan et al. (2016). R was positively related to GPP (path coefficient [PC] = 0.92, Figure 3.5a) and DOC concentrations (PC = 0.05, Figure 3.5a). GPP on the other hand, was positively related to both DOC (PC= 0.26, Figure 3.5a) and TP concentrations (PC = 0.34, Figure 3.5a). TP was positively related to DOC concentration (PC = 0.64, Figure 3.5a) and to the first-axis scores from the PCA of land-use properties (PC = 0.4, Figure 3.5a). DOC concentration was strongly negatively related to d-excess (PC = -

0.71, Figure 3.5a). Different climatic and land use related variables collectively influenced d-excess. Latitude and altitude both were negatively related to d-excess (respectively PC = -0.41, -0.4, Figure 3.5a). d-excess was also negatively related to 30-day prior precipitation (PC = -0.18, Figure 3.5a) and first-axis scores from the PCA of land-use properties (PC = -0.31, Figure 3.5a).

To further understand how DOC concentration and optical or chemical properties of DOM relate to GPP and R we conducted more focused SEMs where we removed the potential effect of climate, land use and lake properties and focused on DOM chemical and optical properties. The SEM based on optical components mediating the relationship between GPP and R had an overall good fit (CFI = 0.96, RMSEA = 0.196, SRMR = 0.084). Both GPP and R were positively related to DOC concentration (respectively PC = 0.65, 0.91, Figure 3.5b), yet GPP was negatively related to %C3 (PC = -0.34, Figure 3.5b), suggesting that this DOM pool tends to suppress GPP, in spite of the overall positive effect of DOM. On the other hand, R was positively related to %C1 (PC = 0.8), suggesting that this may be a DOM pool that tends to enhance R. GPP itself strongly positively influenced %C5 (PC = 7.68, Figure 3.5b), and %C5 in turn positively influenced R (PC = 0.99, Figure 3.5b), whereas %C5 was negatively related to R (PC = -8.14, Figure 3.5b), suggesting that this is a DOM pool that is produced by GPP and consumed by R, therefore representing a direct link between the two metabolic processes.

The SEM associated with the chemical classes of DOM showed a moderate fit based on index values (CFI = 0.89, RMSEA = 0.373, SRMR = 0.109) (Figure 3.5c), and yielded results that were roughly in line with those based on optical properties. DOC concentration again appeared positively related to both GPP (PC = 0.67, Figure 3.5c), and to R (PC = 0.33), and yet GPP appeared to be negatively related to both the High.O.unsat class (in agreement with the negative relationship with %C3 in Figure 3.5b), and also with Aromatic class (respectively PC = -0.41, -0.16, Figure 3.5c), both pools of DOM suppressing GPP. The relationship between GPP and R was modulated by the Aliphatic class. The SEM shows that Aliphatic class was positively related to GPP (PC = 5.71, Figure 3.5c), and as was the case with %C5 in Figure 3.5b, the relationship between Aliphatic class and R was bidirectional, R was positively related to Aliphatic class (PC = 0.47, Figure 3.5c), and Aliphatic class was itself negatively related to R (PC = -5.66, Figure 3.5c), again suggesting that this particular pool is produced by GPP and consumed by R. In contrast to the relationships

with the optical components, the High.O.unsat class was positively related to R (PC = 0.2), suggesting that this pool may fuel some respiration in lakes.

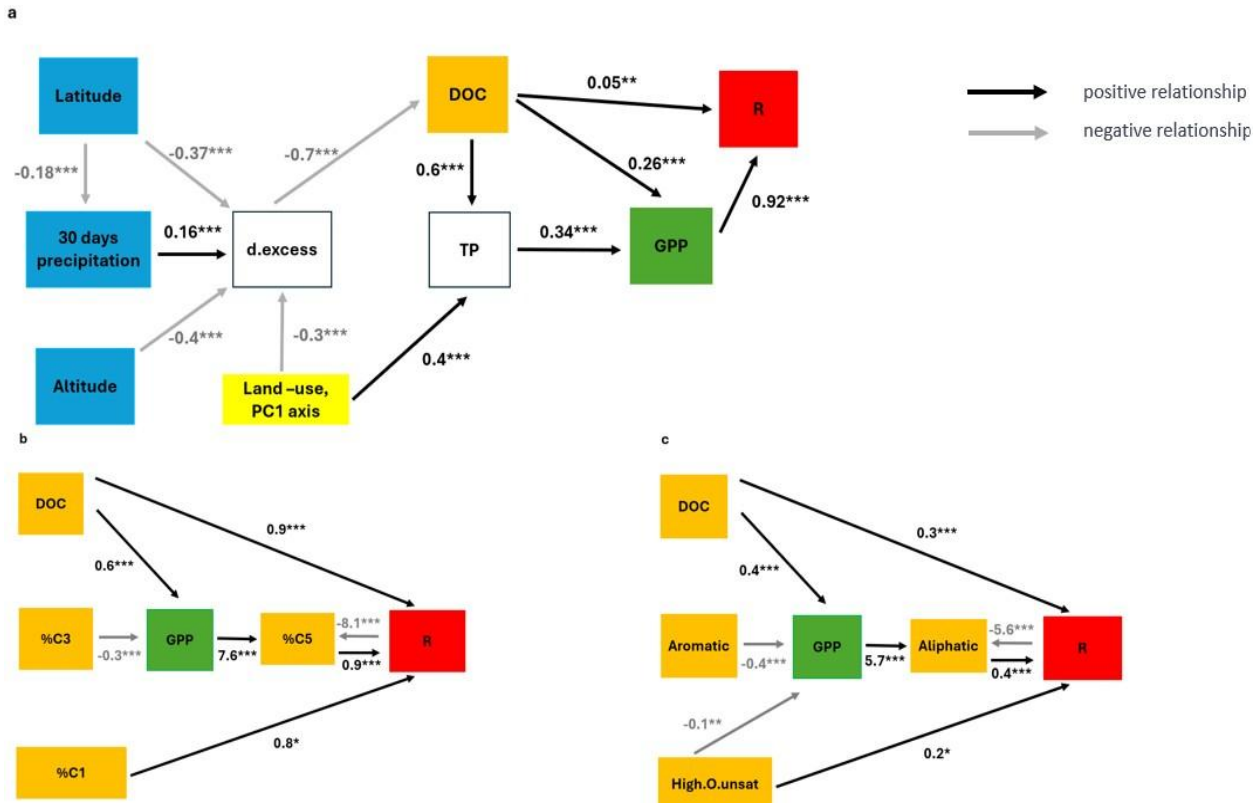

Figure 3.5: Visualization and quantification of the path analysis, a) depicting how landscape and climate variables affect in-lake properties, and, then GPP and R. b) Showing how are the relationship between DOC concentration and its PARAFAC components with GPP and R. c) Showing how are the relationship between DOC concentration and its molecular classes with GPP and R. Path coefficients in gray show a negative relationship and black depicts a positive relationship.

3.5 Discussion

The complex interplay between lake metabolism and DOM composition remains a key uncertainty in understanding C dynamics in inland waters. Our work provides new insights into the drivers of lake metabolism and furthermore explores the potential bidirectional links between ecosystem metabolism and the molecular and optical characteristics of DOM across Canadian lakes. By examining how different metabolic components shape DOM composition, and in turn, how different DOM pools influence metabolic rates, we contribute to a broader understanding of the role of lakes in the global C cycle. Climate and land-use properties directly influenced d-excess, and DOC was negatively related to d-excess.

Previously Herried et al. (2025) also showed similar effects of climate and watershed properties on DOC concentration. Our results reveal a strong positive relationship between lake GPP and both TP and DOC. We also found that GPP and DOC had a strong positive relationship with R (Figure 3.5a). Previous studies have shown that GPP and R vary as a function of DOC (Puts et al. 2022) whereas TP appears to mostly influence GPP (Holgerson et al. 2021). Here we report that (Figure 3.5a) multiple aspects of climate and of land use influence GPP and R via their impact on DOC and nutrients, mostly TP (Figure 3.5a), and that this influence of climate and land use on DOM is mostly exerted via hydrology (d-excess).

Although there have been several studies that have focused on the links between metabolism and DOM, there are very few studies that have actually shown relationships beyond the effect of bulk DOM. We found multiple lines of evidence that autochthonous primary production significantly alters the composition of the lake DOM pool. Our measurements of ambient GPP and R show that there is a very tight coupling between both across Canadian lakes, in agreement with previous studies (Solomon et al. 2013; Oleksy et al. 2021; Klaus et al. 2022; Corman et al. 2023, Ayala-Borda et al. 2024), and we have further shown that GPP influences R but not vice versa. This provides indirect evidence that most of the C fixed within lakes by pelagic (or benthic in the case of shallow lakes) autotrophs is consumed, and suggestive of efficiency internal processing of autochthonous organic matter. We were able to show that the Aliphatic class and the C5 optical component, which are hypothesized to be autochthonous aquatic DOM (de Souza Sierra et al. 1994), are indeed very strongly and positively related to GPP. Protein-like compounds and which make up a portion of Aliphatic class, have been previously associated to primary production in lakes (Liu et al. 2020; Ayala-Borda et al. 2024), and rivers (Hosen et al 2021).

In previous studies we have shown that Aliphatic class proportionately increase along a continuum of water residence time within the aquatic network (Shahabinia et al. submitted, chapter 2), and here we show evidence that this bio-reactive chemical class and its associated optical component are actually being produced and added to DOM pools within lakes and aquatic networks. We also showed a strong coupling between R and Aliphatic class as well as the associated C5 component, suggesting that these are primary substrates for ecosystem respiration (Figure 3.5c). Despite the evidence that Aliphatic are produced within lakes, this molecular class rarely represents more than 10% of the total signal intensity in FT-ICR MS analyses, so this pool does not appear to significantly accumulate within lake DOM, although in previous studies we have shown that it does slightly increase along a gradient of water retention time of lakes (Shahabinia et al. submitted, chapter 2). All points to a relatively small DOM pool that is turning over

extremely rapidly through a tight coupling between primary producers and microbial consumers (Seymour et al., 2017), and which fuels a significant fraction of ecosystem respiration (Hosen et al. 2021; Huryn et al. 2014; Wagner et al. 2017, Ayala-Borda et al. 2024), and this would suggest either shifts in the balance of production vs consumption as a function of water residence time, or more likely, that there is an unreactive component within the Aliphatic pool that tends to persist and accumulate in aquatic DOM. It is interesting to note that the coupling between GPP and R becomes stronger along a gradient of increasing primary production, whereas in unproductive lakes the relationship is much noisier (Figure 3.2b). This would suggest that Aliphatic may be the main substrates of R and mediators of the GPP vs R relationship in eutrophic lakes, whereas other DOM pools may become increasingly important in oligotrophic lakes.

In this regard, we found that two allochthonous DOM pools, High.O.unsat class and optical C1 component were also positively related to R, in agreement with previous studies (Ayala-Borda et al. 2024), suggesting they too partly fuel respiration in lakes, and possibly play a large role in oligotrophic and highly colored systems. It is interesting to note that these same allochthonous DOM pools, Aromatic (C1) and High.O.unsat classes (C3), exert a negative influence on GPP, likely by reducing available light for primary producers. This highlights the fact that the same DOM pools may play multiple and contrasting roles in lake functioning, and therefore in shaping the net metabolic balance of ecosystems.

Whereas our results demonstrated a strong coherence between molecular classes and their associated optical components (Figure 3.4), SEM analyses revealed somewhat different yet complementary linkages of these properties to GPP and R. For example, GPP was negatively related to fluorescent component C3 (Figure 3.5b), and also to its associated molecular class High.O.unsat (Figure 3.5c), yet R was related to the latter, and not to the former (Figure 3.5c). Likewise, R was positively related to C1, and yet Aromatic class (which is associated to C1) had not link to R but was negatively correlated with GPP (Figure 3.5c). These discrepancies could arise from statistical issues, differences in sensitivity (with optical properties potentially capturing certain effects better than molecular data, and vice versa), or the broad categorization of molecular groups. Regardless, it is clear that optical and molecular data provide complementary insights into these relationships.

We have shown that DOM concentration is not the primary driver of lake metabolism; instead, DOM composition plays a more critical role. Some aspects of DOM composition, such as Aliphatic, are likely driven by metabolic processes, while others, like aromatics and highly unsaturated compounds

(High.O.unsat), influence metabolism, though not all major DOM pools are linked to metabolic activity. These unlinked pools neither result from metabolic processes nor exert a measurable effect on them. DOM composition reflects both internal lake dynamics and external influences. For instance, Aliphatics are linked to in-lake C cycling and are not directly connected to land use or climate, whereas other DOM pools are shaped by external factors such as soil and vegetation inputs, wetland presence, runoff patterns, and human activities. As watersheds and climates change, so too does the composition of DOM transported through aquatic networks. These shifts are not merely quantitative (e.g., increased DOM loading) but also qualitative, as seen in widespread lake browning, which alters the characteristics of DOM entering lakes. Different DOM pools respond to distinct environmental and watershed drivers, underscoring the complexity of DOM-metabolism interactions.

3.6 Acknowledgement

We thank the lead PI of NSERC-funded LakePulse Project, Yannick Huot, as well as the many technicians, students, and research professionals involved in the collection and generation of data for the network. Collection and analysis of all variables would have not been possible without the support from members of the CarBBAS team and many technicians, students, and research professionals from various labs across Canada. We are especially grateful to Alice H. Parkes, Katherine Velghe and Marilyne Robidoux for laboratory assistance. We would like to thank the Maxime Fradette for calculating GIS base catchment, land use and climate properties. We would like to thank the Interuniversity Research Group in Limnology (Groupe de recherche interuniversitaire en limnologie [GRIL]) and their funders, the Fonds de recherche—nature et technologie (FRQNT, Québec). This is part of the research program of the NSERC / HydroQuébec Research Chair in Carbon Biogeochemistry in Boreal Aquatic Systems (CarBBas) led by PdG and which also supported AS. JFL was supported by University of Montreal.

3.7 Data availability

All the raw data that supports the findings of this paper will be made entirely available to the public as part of the LakePulse database (https://lakepulse.ca/national-lake-pulsedatabase/).

3.8 Supplementary Information

3.8.1 Fourier transform-ion cyclotron resonance mass spectrometry analysis

The molecular composition of DOM was analyzed using Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry (FT-ICR-MS) on solid-phase extracts obtained from filtered water samples. In FT-ICR-MS, the complex mixture of DOM undergoes ionization and circulates around a large electromagnet which enables precise determination of the masses of numerous molecules. Solid-phase extraction was done using a modified styrene divinyl benzene polymer-type sorbent (100 mg Bond Elut PPL cartridges, Agilent Technologies) (Dittmar et al., 2008). The vials were kept in a dark freezer at -20°C until further analysis. The DOM solid-phase extracts were analyzed using a 7T Bruker SolariX XR (FT-ICR-MS) at the Water Quality Centre, Trent University, Peterborough, Canada. In addition, each day external calibration was performed using electrospray ionization (ESI) and sodium trifluoroacetate (NaTFA) in methanol at a concentration of 0.1 mg/mL, in both positive and negative ion mode. Data acquisition was conducted using the Bruker ftms Control software (version 2.1.0), covering a mass range spanning from m/z 200 to 1000. The data was exported using the Bruker Compass DataAnalysis software (version 5.0). The samples were injected individually at a flow rate of 120 µL/h to acquire 300 spectra scans. We processed the resulting masses for quality and assigned molecular formulae following established protocols from the ICBM-OCEAN tool, an open platform for DOM mass spectra processing (Merder et al., 2020). Molecular formulas were assigned based on the following elemental constraints: C_{4-100}, H_{4-200}, O_{1-70}, N_{0-4}, S_{0-2}, and P_{0-1}. Furthermore, we cleaned the dataset, samples with sum intensities below 50 were removed, and the intensities of isomers (compounds with the same molecular formula but distinct atom arrangements) were summed.

3.8.2 Oxygen isotope method

We estimated summertime mixed-layer metabolic rates in lakes using an oxygen isotope ($\delta^{18}O_2$) approach, which combines ambient O_2 concentration and its isotopic signature, as detailed in Bogard et al. (2017) and Bocaniov et al. (2012, 2015). This method was chosen because it offers a straightforward and efficient way to provide an ecosystem-level estimate of metabolism, where samples could be collected within a single day. Other methods, which require extended incubation times or intensive sampling efforts, were not feasible for this large-scale study. This method accounts for the effects of air-lake gas exchange on the DO pool and due to its free-water nature, it does not distinguish between biotic and abiotic DO fluxes,

therefore, the overall R rates presented here include several O_2 consumption pathways, such as pelagic R, organic matter photo-oxidation, and benthic metabolism (portion of sediments in contact with mixed layer). This method assumes steady state conditions, where there is no net change in either DO concentration or $\delta^{18}O_2$ over a diel cycle. As proposed by Bogard et al. (2017), we estimated volumetric GPP and R with the following equations:

$$GPP = \left(\frac{kO2}{Zmix}\right) x \frac{[(DO\ x\ (b-c)) - (DOsat\ x\ (a-c))]}{d-c} \quad (1)$$

$$R = \left(\frac{kO2}{Zmix}\right) x \frac{[(DO\ x\ (b-d)) - (DOsat\ x\ (a-d))]}{d-c} \quad (2)$$

where Z_{mix} is the mixed layer depth, DO and DO_{sat} are the ambient and saturation DO (mg L^{-1}) concentrations, respectively, and k_{O2} (m d^{-1}) is the gas transfer coefficient of oxygen , a = AF_{atm} x a_s x a_g, b = AF_{DO} x a_g, c = AF_{DO} x a_c and d = AF_{H2O} x a_p. We considered fractionation factors for gas exchange at the air-water interface (a_g = 0.9972, Knox et al. 1992), gas solubility in water (a_s = 1.0007, Benson & Krause, 1980), and DO production through photosynthesis (a_p = 1.000, Guy et al. 1993) and consumption of DO at ecosystem-level (a_c = 0.985, Bogard et al. 2017). We calculated the atomic fractions for dissolved oxygen (AF_{DO}), water (AF_{H2O}), and atmosphere (AF_{atm}) following Bogard et al. (2020), based on Hotchkiss and Hall (2014) as in Equation 3, where i is the atomic fraction:

$$AFi = \frac{Rsample}{1+Rsample} \quad (3)$$

Where AF_{DO}, AF_{H2O}, and AF_{atm} represent the atomic fraction of ^{18}O in DO, water, and atmospheric O_2, respectively. We also estimated NEP as the difference between GPP and R, and calculated the ratios of GPP:R.

3.8.3 Figures and tables

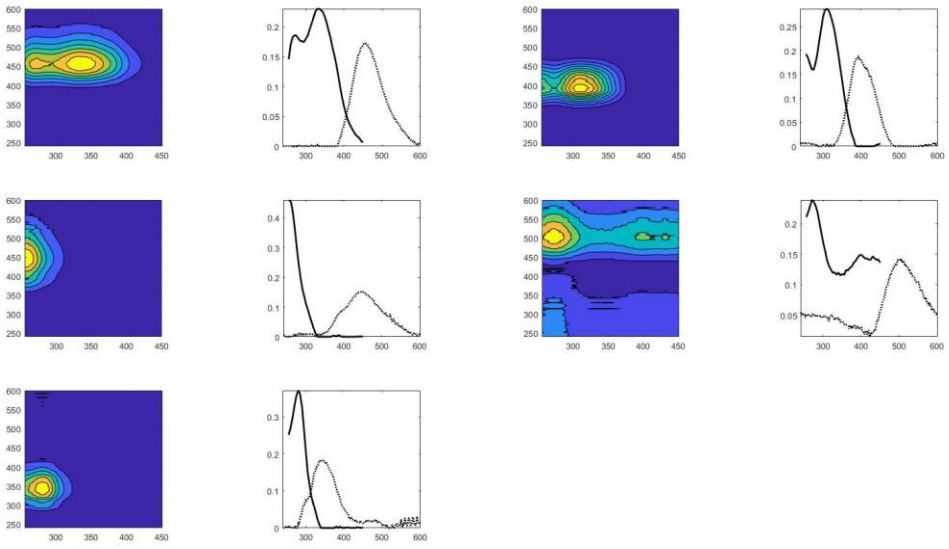

Figure S3.1: Fluorescence signature of the components identified by the PARAFAC model.

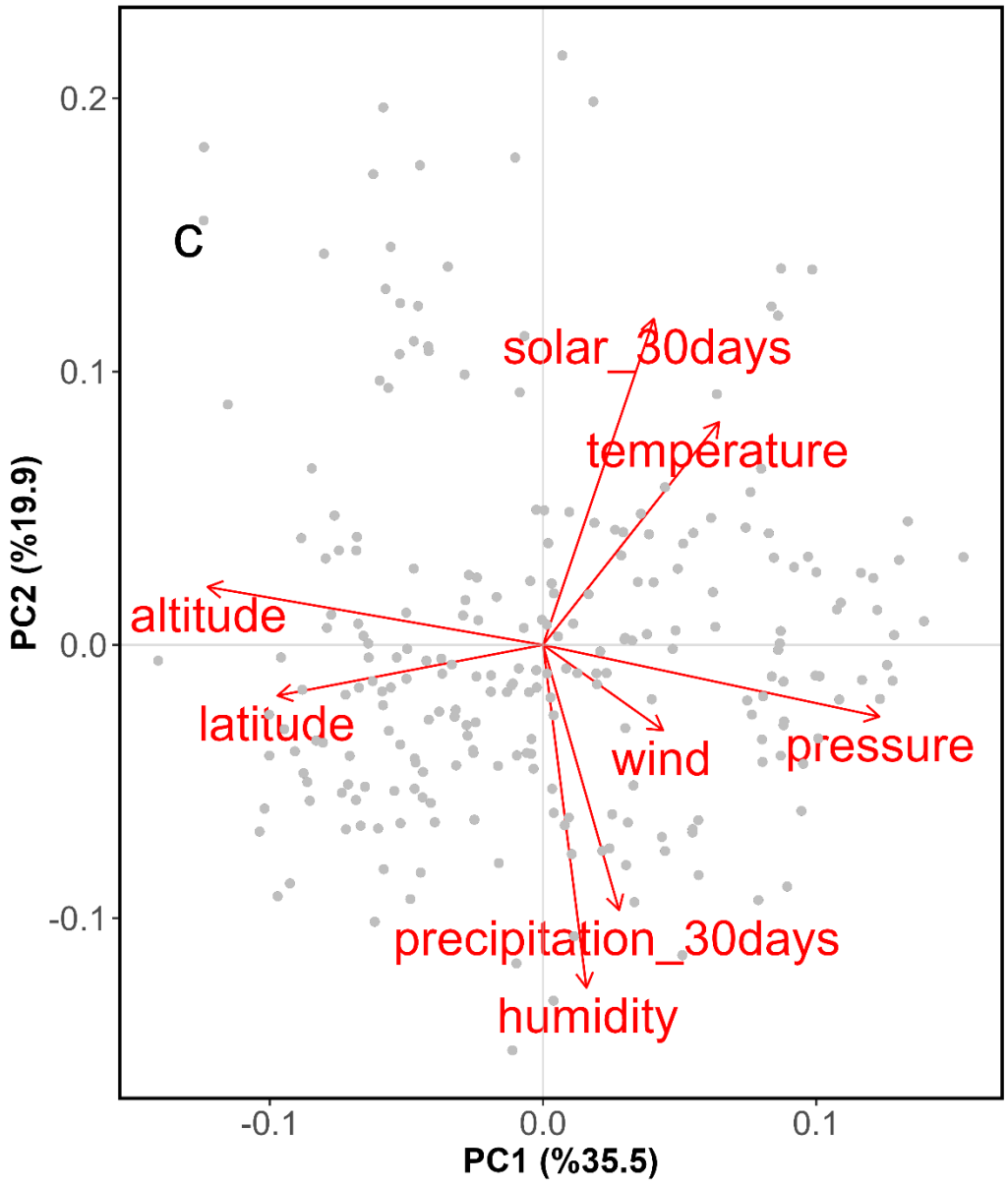

Figure S3.2: Principal component analysis (PCA) on climate properties included in this study.

CONCLUSION

4.1 General conclusion

The overall aim of this thesis was to increase our understanding of the functioning of northern lakes by examining summertime metabolism and DOM concentration and composition, which together play a crucial role in shaping ecosystem processes and C dynamics. This thesis contributed to our understanding of patterns and drivers of lake metabolism and DOM across large spatial scales. To do so, I measured metabolism using the oxygen isotope method, which allowed me to estimate ecosystem-scale metabolism from a single point sample. I quantified the optical component of DOM composition using the PARAFAC method and studied molecular properties of DOM using the FT-ICR MS method. I investigated local and regional drivers and explored the interaction between these drivers of lake metabolism and DOM composition at an unprecedented continental scale. I identified complex links between different components of metabolism and different optical and molecular pools of DOM. This work contributed to an understanding of the role of freshwater ecosystems in the global C cycle by providing a comprehensive dataset for the literature, including measurements of GPP, R, and their balance, as well as optical and molecular properties of DOM. The thesis advanced our understanding of lake functioning by addressing important gaps in current knowledge, as I detail below.

First, I showed that GPP and R varied by four orders of magnitude and that there was strong coupling between GPP and R in lakes across Canada. My results showed that GPP is primarily driven by nutrient availability, with total phosphorus, total nitrogen, and DOC playing important roles. Although R responded positively to the same drivers as GPP, the strength of these relationships was not the same for GPP and R. This decoupled GPP from R, especially in unproductive oligotrophic lakes, and this is reflected in patterns in NEP (Chapter 1). The coupling between GPP and R became stronger along a gradient of increasing primary production, suggesting that the Aliphatic class may be the main substrate of R and mediators of the GPP vs. R relationship in eutrophic lakes, whereas other DOM pools, such as the Aromatic class may be more important in oligotrophic lakes (Chapter 3).

Second, I identified the drivers of lake metabolism (e.g., relationships between metabolism and nutrients or *Chla*), and these drivers were comparable to those reported in different studies, except for the observed

positive relationship between GPP and DOC—a pattern that contrasts with widely held theoretical expectations of a unimodal (bell-shaped) GPP–DOC relationship. In my study, I demonstrated that NEP is highly sensitive to light and CDOM, with light enhancing GPP while CDOM reduces light availability and may also increase R. This highlighted the complex role that DOC (and the associated CDOM) plays in these systems (Chapter 1). Across all lakes, DOC and CDOM concentrations were strongly correlated (Chapter 2), suggesting dominance of terrestrial inputs across Canadian lakes. The strong positive relationships between DOC and both GPP and R indicate that allochthonous organic matter inputs significantly influence lake metabolism across lakes by delivering nutrients but also by acting as a substrate for R. Yet the enhancement of R by DOC and the depression of GPP by CDOM appear to exceed the positive effect of DOC on GPP, leading to reduced NEP. As a result, the most negative NEP values I observed were in highly colored, DOC-rich lakes, likely reflecting this overall balance of effects of allochthonous organic matter on lake metabolism (Chapter 1). In this regard, I showed that DOC concentration is not the only major factor influencing lake metabolism, rather DOM composition plays a central role. I showed that Aromatic and High.O.unsat classes were positively related to R but negatively related to GPP, suggesting that the same DOM pools may play multiple and contrasting roles in lake functioning (Chapter 3).

In Chapter 2, I showed that at a continental scale, d-excess and TP concentration, as indicators of hydrology and trophic status, shape both DOC concentration and the optical and molecular composition of DOM. I showed that d-excess plays the leading role in determining DOM concentration and chemical composition. I demonstrated that some DOM pools, such as the Aliphatic class and the associated C5 optical component, tend to become enriched along the hydrological continuum (Chapter 2) and I further demonstrated that these chemical components are actually driven by primary production within the aquatic network (Chapter 3). On the other hand, other DOM pools, such as the Aromatic class and C1 and C3 optical components, tend to decrease along a gradient of water retention time within the aquatic network, likely due to photochemical removal (Chapter 2) and I further demonstrated that these components are not influenced by GPP or R, suggesting that other external factors must be involved in driving these components (Chapter 3).

I showed that PARAFAC component C1 was strongly related to the Aromatic class, component C5 was strongly aligned with the Aliphatic class, and component C3 was strongly coupled to the High.O.unsat class, suggesting that these optical indices reflect the underlying molecular composition (Chapters 2 and 3). Yet I did not observe the same links between them and metabolic rates, suggesting that these optical and

molecular components provide complementary information on the links between DOM and metabolism (Chapter 3).

In Chapters 1 and 2, I bridged the gap between previous local and global studies and revealed regional differences in the relationships between GPP, R, and NEP and their environmental drivers. I also found regional variation in the associations between DOM molecular classes and d-excess. By sampling lakes across a broad spatial scale, I was able to disentangle the complex interplay between local and regional variables. I showed that patterns in lake metabolism at the continental scale are the result of the interaction between local drivers (i.e., TP, TN, DOC), which regulate metabolism at individual lakes, and landscape-level drivers (i.e., altitude, d-excess), which integrate regional properties and modulate the collective response of lakes to the environment and climate. These findings would not have been possible to detect if I had only studied lake metabolism at a local scale (Chapter 1). In Chapter 2, I observed regional differences in the associations between DOM molecular classes and d-excess, suggesting that the extent of evaporation and the associated water residence time varies greatly between basins. This enabled me to reconstruct the source DOM. The reconstructed composition of source DOM derived from the regional models was related to regional watershed features, including average regional topography, soil properties, and climate. I revealed that the reconstructed source DOM composition, together with the extent of its processing along the network, shapes the DOM composition in lakes across Canada. Yet my results suggest that the latter plays the leading role in determining cross-regional patterns in DOM composition.

One of the major challenges in linking lake metabolism and GHG emission is related to determining the net metabolic balance. This is due to numerous methodological discrepancies that complicate interpretation of NEP and GPP:R in lakes. In Chapter 1, using a consistent and ecosystem-scale method, I determined NEP and the GPP:R ratio at the continental scale, and I showed that the proportion of lakes that are heterotrophic is around 50%, with average regional NEP in some basins being net autotrophic (+NEP) and net heterotrophic (–NEP) in others. My results showed that the balance between primary production and respiration appears to have a regional structure, yet with considerable variation of NEP within a given region. This finding has consequences for various aspects of lake functioning, and one might hypothesize that, as a whole, these lakes neither emit nor take up significant amounts of CO_2, and that they do not store significant amounts of organic C. However, it is possible that lakes that have large positive NEP may accumulate proportionally more organic carbon than other lakes, because these lakes may indeed have a greater autochthonous source of sedimenting matter. This might explain some of the inter-

lake and cross-regional differences that have been observed in lake C burial (Heathcote et al. 2015). This implies that regions where lakes have consistently positive NEP should, on average, have higher rates of organic C burial than regions where lakes have consistently negative NEP, but this hypothesis remains to be tested.

REFERENCES

Abbasi, M., M. Peacock, S. Drakare, J. Hawkes, E. Jakobsson, and D. Kothawala. 2024. Water residence time is an important predictor of dissolved organic matter composition and drinking water treatability. Water Res. 260: 121910. http://doi:10.1016/j.watres.2024.121910

Anderson, N. J., and C. A. Stedmon. 2007. The effect of evapoconcentration on dissolved organic carbon concentration and quality in lakes of SW Greenland. Freshwater Biol. 52: 280–289. http://doi:10.1111/j.1365-2427.2006.01688.x

Ask, J., J. Karlsson, and M. Jansson. 2012. Net ecosystem production in clear-water and brown-water lakes. Global Biogeochem. Cycles 26. http://doi:10.1029/2010GB003951

ASTM. 2003. Standard tables for reference solar spectral irradiances: Direct normal and hemispherical on 37° tilted surface. *ASTM International Standard G173-03*.

Ayala-Borda, P., M. J. Bogard, G. Grosbois, V. Prėskienis, J. M. Culp, M. Power, and M. Rautio. 2024. Dominance of net autotrophy in arid landscape low relief polar lakes, Nunavut, Canada. Glob. Change Biol. 30, e17193. http://doi:10.1111/gcb.17193

Bailey, V. L., A. P. Smith, M. Tfaily, S. J. Fansler, and B. Bond-Lamberty. 2017. Differences in soluble organic carbon chemistry in pore waters sampled from different pore size domains. Soil Biol. Biochem. 107: 133–143. http://doi:10.1016/j.soilbio.2016.11.025

Barbosa, P. M., P. Bodmer, M. Stadler, F. Rust, A. Tremblay, and P. A. del Giorgio. 2023. Ecosystem metabolism is the dominant source of carbon dioxide in three young boreal cascade-reservoirs (La Romaine Complex, Québec). J. Geophys. Res. Biogeosci. 128. http://doi:10.1029/2022JG007253

Barnes, D. J. 1983. Profiling coral reef productivity and calcification using pH and oxygen electrodes. J. Exp. Mar. Biol. Ecol. 71: 73–89. http://doi:10.1016/0022-0981(83)90036-9

Bartlett, J. S., Á. M. Ciotti, R. F. Davis, and J. J. Cullen. 1998. The spectral effects of clouds on solar irradiance. J. Geophys. Res. Oceans 103: 31017–31031. http://doi:10.1029/1998JC900002

Barth, J. A. C., T. Andrew, and B. Mike. 2004. Automated analyses of 18O/16O ratios in dissolved oxygen from 12-mL water samples. Limnol. Oceanogr. Methods 2: 35–42. http://doi:10.4319/lom.2004.2.35

Battin, T., S. Luyssaert, L. Kaplan, and others. 2009. The boundless carbon cycle. Nat. Geosci. 2: 598–600. http://doi:10.1038/ngeo618

Behnke, M. I., J. W. McClelland, S. E. Tank, A. M. Kellerman, R. M. Holmes, N. Haghipour, and others. 2021. Pan-Arctic Riverine Dissolved Organic Matter: Synchronous Molecular Stability, Shifting Sources and Subsidies. Glob. Biogeochem. Cycles 35. http://doi:10.1029/2020GB006871

Bender, M., K. Grande, K. Johnson, and others. 1987. A comparison of four methods for determining planktonic community production. Limnol. Oceanogr. 32: 1085–1098. http://doi:10.4319/lo.1987.32.5.1085

Bocaniov, S. A., S. L. Schiff, and R. E. H. Smith. 2015. Non steady-state dynamics of stable oxygen isotopes for estimates of metabolic balance in large lakes. J. Great Lakes Res. 41: 719–729. http://doi:10.1016/j.jglr.2015.05.013

Bocaniov, S. A., S. L. Schiff, and R. E. H. Smith. 2012. Plankton metabolism and physical forcing in a productive embayment of a large oligotrophic lake: Insights from stable oxygen isotopes. Freshwater Biol. 57: 481–496. http://doi:10.1111/j.1365-2427.2011.02715.x

Bogard, M. J., N. F. St-Gelais, D. Vachon, and P. A. del Giorgio. 2020. Patterns of spring/summer open-water metabolism across boreal lakes. Ecosystems. http://doi:10.1007/s10021-020-00487-7

Bogard, M. J., D. Vachon, N. F. St.-Gelais, and P. A. del Giorgio. 2017. Using oxygen stable isotopes to quantify ecosystem metabolism in northern lakes. Biogeochemistry 133: 347–364. http://doi:10.1007/s10533-017-0338-5

Bogard, M. J., and P. A. del Giorgio. 2016. The role of metabolism in modulating CO_2 fluxes in boreal lakes. Glob. Biogeochem. Cycles 30: 1509–1525. http://doi:10.1002/2016GB005463

Butturini, A., A. Guarch, A. M. Romaní, A. Freixa, S. Amalfitano, S. Fazi, and E. Ejarque. 2016. Hydrological conditions control in situ DOM retention and release along a mediterranean river. Water Res. 99: 33–45. http://doi:10.1016/j.watres.2016.04.036

Caffrey, J. M. 2004. Factors controlling net ecosystem metabolism in U.S. estuaries. Estuaries 27: 56–69. http://doi:10.1007/BF02803563

Casas-Ruiz, J. P., N. Catalán, L. Gómez-Gener, D. von Schiller, B. Obrador, D. N. Kothawala, P. López, S. Sabater, and R. Marcé. 2017. A tale of pipes and reactors: Controls on the in-stream dynamics of dissolved organic matter in rivers. Limnol. Oceanogr. 62: S85–S94. http://doi:10.1002/lno.10471

Catalán, N., R. Marcé, D. N. Kothawala, and L. J. Tranvik. 2016. Organic carbon decomposition rates controlled by water retention time across inland waters. Nat. Geosci. 9: 501–504. http://doi:10.1038/ngeo2720

Coble, P. G. 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Mar. Chem. 51: 325–346. http://doi:10.1016/0304-4203(95)00062-3

Coble, P. G., S. A. Green, N. V. Blough, and R. B. Gagosian. 1990. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. Nature 348: 432–435. http://doi:10.1038/348432a0

Cole, J. J., Y. T. Prairie, N. F. Caraco, and others. 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems 10: 171–184. http://doi:10.1007/s10021-006-9013-8

Cooper, W. T., J. C. Chanton, J. D'Andrilli, S. B. Hodgkins, D. C. Podgorski, A. C. Stenson, and others. 2022. A History of Molecular Level Analysis of Natural Organic Matter by FTICR Mass Spectrometry and The Paradigm Shift in Organic Geochemistry. Mass Spectrom. Rev. 41: 215–239. http://doi:10.1002/mas.21663

Cory, R. M., and G. W. Kling. 2018. Interactions between sunlight and microorganisms influence dissolved organic matter degradation along the aquatic continuum. Limnol. Oceanogr. Lett. 3: 102–116. http://doi:10.1002/lol2.10060

Corman, J. R., J. A. Zwart, J. Klug, and others. 2023. Response of lake metabolism to catchment inputs inferred using high-frequency lake and stream data from across the northern hemisphere. Limnol. Oceanogr. 68: 2617–2631. http://doi:10.1002/lno.12449

Creed, I. F., F. D. Beall, T. A. Clair, P. J. Dillon, and R. H. Hesslein. 2008. Predicting export of dissolved organic carbon from forested catchments in glaciated landscapes with shallow soils. Glob. Biogeochem. Cycles 22. http://doi:10.1029/2008GB003294

Cuthbert, I. D., and P. del Giorgio. 1992. Toward a standard method of measuring color in freshwater. Limnol. Oceanogr. 37: 1319–1326. http://doi:10.4319/lo.1992.37.6.1319

Dansgaard, W. 1964. Stable isotopes in precipitation. Tellus A: Dyn. Meteorol. Oceanogr. 16: 436. http://doi:10.3402/tellusa.v16i4.8993

Dee, S., A. Bailey, J. L. Conroy, A. Atwood, S. Stevenson, J. Nusbaumer, and D. Noone. 2023. Water isotopes, climate variability, and the hydrological cycle: recent advances and new frontiers. Environ. Res. Clim. 2: 022002. http://doi:10.1088/2752-5295/accbe1

Deemer, B. R., J. A. Harrison, S. Li, and others. 2016. Greenhouse Gas Emissions from Reservoir Water Surfaces: A New Global Synthesis. Bioscience 66: 949–964. http://doi:10.1093/biosci/biw117

De Haan, H. 1993. Solar UV‐light penetration and photodegradation of humic substances in peaty lake water. Limnol. Oceanogr. 38: 1072–1076. http://doi:10.4319/lo.1993.38.5.1072

de Souza Sierra, M. M., O. F. X. Donard, M. Lamotte, C. Belin, and M. Ewald. 1994. Fluorescence spectroscopy of coastal and marine waters. Mar. Chem. 47: 127-144. http://doi:10.1016/0304-4203(94)90012-4

del Giorgio, P. A., J. J. Cole, N. F. Caraco, and R. H. Peters. 1999. Linking planktonic biomass and metabolism to net gas fluxes in northern temperate lakes. Ecology 80: 1422–1431. http://doi:10.1890/0012-9658(1999)080[1422:LPBAMT]2.0.CO;2

del Giorgio, P. A., and R. H. Peters. 1994. Patterns in planktonic P:R ratios in lakes: Influence of lake trophy and dissolved organic carbon. Limnol. Oceanogr. 39: 772–787. http://doi:10.4319/lo.1994.39.4.0772

del Giorgio, P. A., and R. H. Peters. 1993. Balance between phytoplankton production and plankton respiration in lakes. Can. J. Fish. Aquat. Sci. 50. http://doi:10.1139/f93-032

Deutsch, E. S., M. J. Fortin, and J. A. Cardille. 2022. Assessing the current water clarity status of ~100,000 lakes across southern Canada: A remote sensing approach. Sci. Total Environ. 826: 153971. http://doi:10.1016/j.scitotenv.2022.153971

Dittmar, T., B. Koch, N. Hertkorn, and G. Kattner. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnol. Oceanogr. Methods 6: 230–235. http://doi:10.4319/lom.2008.6.230

Dupont, A., M. Botrel, N. Fortin St-Gelais, T. Poisot, and R. Maranger. 2023. A social–ecological geography of southern Canadian lakes. Facets 8: 1–16. http://doi:10.1139/facets-2023-0025

Falkowski, P., R. J. Scholes, E. Boyle, and others. 2000. The global carbon cycle: A test of our knowledge of earth as a system. Science 290: 291–296. http://doi:10.1126/science.290.5490.291

Fan, Y., J. Chen, G. Shirkey, R. John, S. R. Wu, H. Park, and C. Shao. 2016. Applications of structural equation modeling (SEM) in ecological studies: An updated review. Ecol. Process. 5. http://doi:10.1186/s13717-016-0063-3

Fasching, C., A. J. Ulseth, J. Schelker, G. Steniczka, and T. J. Battin. 2016. Hydrology controls dissolved organic matter export and composition in an alpine stream and its hyporheic zone. Limnol. Oceanogr. 61: 558–571. http://doi:10.1002/lno.10232

Fee, E. J., R. E. Hecky, S. E. M. Kasian, and D. R. Cruikshank. 1996. Effects of lake size, water clarity, and climatic variability on mixing depths in Canadian Shield lakes. Limnol. Oceanogr. 41: 912–920. http://doi:10.4319/lo.1996.41.5.0912

Forget, M.-H., R. Carignan, and C. Hudon. 2009. Influence of diel cycles of respiration, chlorophyll, and photosynthetic parameters on the summer metabolic balance of temperate lakes and rivers. Can. J. Fish. Aquat. Sci. http://doi:10.1139/f09-058

Galewsky, J., H. C. Steen-Larsen, R. D. Field, J. Worden, C. Risi, and M. Schneider. 2016. Stable isotopes in atmospheric water vapor and applications to the hydrologic cycle. Rev. Geophys. 54: 809–865. http://doi:10.1002/2015RG000512

Gattuso, J. P., M. Pichon, B. Delesalle, and M. Frankignoulle. 1993. Community metabolism and air-sea CO_2 fluxes in a coral reef ecosystem (Moorea, French Polynesia). Mar. Ecol. Prog. Ser. 96: 259–267. http://doi:10.3354/meps096259

Gazeau, F., A. V. Borges, C. Barrón, and others. 2005. Net ecosystem metabolism in a micro-tidal estuary (Randers Fjord, Denmark): Evaluation of methods. Mar. Ecol. Prog. Ser. 301: 23–41. http://doi:10.3354/meps301023

Grey, J., R. I. Jones, and D. Sleep. 2001. Seasonal changes in the importance of the source of organic matter to the diet of zooplankton in Loch Ness, as indicated by stable isotope analysis. Limnol. Oceanogr. 46: 505–513. http:// doi:10.4319/lo.2001.46.3.0505

Gudasz, C., M. Ruppenthal, K. Kalbitz, C. Cerli, S. Fiedler, Y. Oelmann, A. Andersson, and J. Karlsson. 2017. Contributions of terrestrial organic carbon to northern lake sediments. Limnol. Oceanogr. Lett. 2: 218–227. http://doi:10.1002/lol2.10051

Guy, R. D., M. L. Fogel, and J. A. Berry. 1993. Photosynthetic fractionation of the stable isotopes of oxygen and carbon. Plant Physiol. 101: 37-47. http://doi: 10.1104/pp.101.1.37

Hanson, P. C., S. R. Carpenter, N. Kimura, C. Wu, S. P. Cornelius, and T. K. Kratz. 2008. Evaluation of metabolism models for free-water dissolved oxygen methods in lakes. Limnol. Oceanogr. Methods 6: 454–465. http://doi:10.4319/lom.2008.6.454

Hanson, P. C., D. L. Bade, S. R. Carpenter, and T. K. Kratz. 2003. Lake metabolism: Relationships with dissolved organic carbon and phosphorus. Limnol. Oceanogr. 48: 1112–1119. http://doi:10.4319/lo.2003.48.3.1112

Hawkes, J. A., J. D'Andrilli, J. N. Agar, M. P. Barrow, S. M. Berg, N. Catalán, and others. 2020. An international laboratory comparison of dissolved organic matter composition by high resolution mass spectrometry: Are we getting the same answer? Limnol. Oceanogr. Methods 18: 235–258. http://doi:10.1002/lom3.10364

Heathcote, A. J., N. J. Anderson, Y. T. Prairie, D. R. Engstrom, and P. A. del Giorgio. 2015. Large increases in carbon burial in northern lakes during the Anthropocene. Nat. Commun. 6. http://doi:10.1038/ncomms10016

Helms, J. R., D. J. Kieber, E. C. Minor, J. D. Ritchie, A. Stubbins, and K. Mopper. 2008. Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. Limnol. Oceanogr. 53: 955–969. http://doi:10.4319/lo.2008.53.3.0955

Herreid, A. M., H. M. Fazekas, S. J. Nelson, A. S. Wymore, D. Murray, R. K. Varner, and W. H. McDowell. 2025. Climate displaces deposition as dominant driver of dissolved organic carbon concentrations in historically acidified lakes. Biogeochemistry 168. http://doi:10.1007/s10533-024-01193-5

Hessen, D. O. 1992. Dissolved organic carbon in a humic lake: Effects on bacterial production and respiration. Hydrobiologia 229: 115–123. http://doi:10.1007/BF00006995

Hoellein, T. J., D. A. Bruesewitz, and D. C. Richardson. 2013. Revisiting Odum (1956): A synthesis of aquatic ecosystem metabolism. Limnol. Oceanogr. 58: 2089–2100. http://doi:10.4319/lo.2013.58.6.2089

Holgerson, M. A., R. A. Hovel, P. T. Kelly, L. E. Bortolotti, J. A. Brentrup, A. R. Bellamy, S. K. Oliver, and A. J. Reisinger. 2022. Integrating ecosystem metabolism and consumer allochthony reveals nonlinear drivers in lake organic matter processing. Limnol. Oceanogr. 67: S71–S85. http://doi:10.1002/lno.11907

Hosen, J. D., K. S. Aho, J. H. Fair, E. D. Kyzivat, S. Matt, J. Morrison, and others. 2021. Source Switching Maintains Dissolved Organic Matter Chemostasis Across Discharge Levels in a Large Temperate River Network. Ecosystems 24: 227–247. http://doi:10.1007/s10021-020-00514-7

Huguet, A., L. Vacher, S. Relexans, S. Saubusse, J. M. Froidefond, and E. Parlanti. 2009. Properties of fluorescent dissolved organic matter in the Gironde Estuary. Org. Geochem. 40: 706-719. doi:10.1016/j.orggeochem.2009.03.002

Huot, Y., C. A. Brown, G. Potvin, and others. 2019. The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing science for water management in a changing climate. Sci. Total Environ. 695. http://doi:10.1016/j.scitotenv.2019.133668

Huryn, A. D., J. P. Benstead, and S. M. Parker. 2014. Seasonal changes in light availability modify the temperature dependence of ecosystem metabolism in an arctic stream. Ecology 95: 2826–2839. http://doi:10.1890/13-1903.1

Hutchins, R. H., P. Aukes, S. L. Schiff, T. Dittmar, Y. T. Prairie, and P. A. del Giorgio. 2017. The optical, chemical, and molecular dissolved organic matter succession along a boreal soil-stream-river continuum. J. Geophys. Res. Biogeosci. 122: 2892–2908. http://doi:10.1002/2017JG004094

Ines Fung, J.-E. L. 2008. "Amount effect" of water isotopes and quantitative analysis of post-condensation processes. Hydrol. Process. 22: 2267–2274. http://doi:10.1002/hyp.6637

JMP®, Version Pro 16.0.0. 2023. SAS Institute Inc., Cary, NC.

Johnston, S. E., R. G. Striegl, M. J. Bogard, M. M. Dornblaser, D. E. Butman, A. M. Kellerman, and others. 2020. Hydrologic connectivity determines dissolved organic matter biogeochemistry in northern high-latitude lakes. Limnol. Oceanogr. 1–17. http://doi:10.1002/lno.11417

Johnston, S. E., M. J. Bogard, J. A. Rogers, and others. 2019. Constraining dissolved organic matter sources and temporal variability in a model sub-Arctic lake. Biogeochemistry 146: 271–292. http://doi:10.1007/s10533-019-00619-9

Jones, T. G., Evans, C. D., Jones, D. L., Hill, P. W. and Freeman, C. (2016). Transformations in DOC along a source to sea continuum ; impacts of photo-degradation , biological processes and mixing. Aquatic Sciences, 78. http://doi.org/10.1007/s00027-015-0461-0

Jones, J. R., J. L. Graham, D. V. Obrecht, J. D. Harlan, M. F. Knowlton, C. Pollard, and others. 2024. Role of edaphic, hydrologic, and land cover variables in determining dissolved organic carbon in Missouri (USA) reservoirs and streams. Lake Reserv. Manage. 40: 177–195. http://doi:10.1080/10402381.2024.2326057

Karlsson, J., P. Byström, J. Ask, P. Ask, L. Persson, and M. Jansson. 2009. Light limitation of nutrient-poor lake ecosystems. Nature 460: 506–509. http://doi:10.1038/nature08179

Karlsson, J., A. Jonsson, and M. Jansson. 2005. Productivity of high-latitude lakes: Climate effect inferred from altitude gradient. Glob. Change Biol. 11: 710–715. http://doi:10.1111/j.1365-2486.2005.00945.x

Kellerman, A. M., A. Arellano, D. C. Podgorski, E. E. Martin, J. B. Martin, K. M. Deuerling, and others. 2020. Fundamental drivers of dissolved organic matter composition across an Arctic effective precipitation gradient. Limnol. Oceanogr. 65: 1217–1234. http://doi:10.1002/lno.11385

Kellerman, A. M., F. Guillemette, D. C. Podgorski, G. R. Aiken, K. D. Butler, and R. G. M. Spencer. 2018. Unifying concepts linking dissolved organic matter composition to persistence in aquatic ecosystems. Environ. Sci. Technol. 52: 2538–2548. http://doi:10.1021/acs.est.7b05513

Kellerman, A. M., T. Dittmar, D. N. Kothawala, and L. J. Tranvik. 2014. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. Nat. Commun. 5: 1–8. http://doi:10.1038/ncomms4804

Kelly, P. T., C. T. Solomon, J. A. Zwart, and S. E. Jones. 2018. A framework for understanding variation in pelagic gross primary production of lake ecosystems. Ecosystems 21: 1364–1376. http://doi:10.1007/s10021-018-0226-4

Klaus, M., H. A. Verheijen, J. Karlsson, and D. A. Seekell. 2022. Depth and basin shape constrain ecosystem metabolism in lakes dominated by benthic primary producers. Limnol. Oceanogr. 67: 2763–2778. http://doi:10.1002/lno.12236

Knox, M., P. D. Quay, and D. Wilbur. 1992. Kinetic isotopic fractionation during air-water gas transfer of O_2, N_2, CH_4 and H_2. J. Geophys. Res. 97: 335-343. http://doi: 10.1029/92JC00949

Koch, B. P., and T, Dittmar. 2006. From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. Rapid Communications in Mass Spectrometry, 20, 926–932. http://doi.org/10.1002/rcm.2386

Kothawala, D. N., X. Ji, H. Laudon, A. M. Ågren, M. N. Futter, S. J. Köhler, and L. J. Tranvik. 2015. The relative influence of land cover, hydrology, and in-stream processing on the composition of dissolved organic matter in boreal streams. J. Geophys. Res. Biogeosci. 120: 1491–1505. http://doi:10.1002/2015JG002946

Kothawala, D. N., C. A. Stedmon, R. A. Müller, G. A. Weyhenmeyer, S. J. Köhler, and L. J. Tranvik. 2014. Controls of dissolved organic matter quality: Evidence from a large-scale boreal lake survey. Glob. Change Biol. 20: 1101–1114. http://doi:10.1111/gcb.12488

Kothawala, D. N., E. von Wachenfeldt, B. Koehler, and L. J. Tranvik. 2012. Selective loss and preservation of lake water dissolved organic matter fluorescence during long-term dark incubations. Sci. Total Environ. 433: 238–246. http://doi:10.1016/j.scitotenv.2012.06.029

Kurek, M. R., K. P. Wickland, N. A. Nichols, A. M. McKenna, S. M. Anderson, M. M. Dornblaser, and others. 2024. Linking Dissolved Organic Matter Composition to Landscape Properties in Wetlands Across the United States of America. Glob. Biogeochem. Cycles 38: 1–21. http://doi:10.1029/2023GB007917

Kurek, M. R., F. Garcia-Tigreros, K. P. Wickland, K. E. Frey, M. M. Dornblaser, R. G. Striegl, and others. 2023. Hydrologic and Landscape Controls on Dissolved Organic Matter Composition Across Western North American Arctic Lakes. Glob. Biogeochem. Cycles 37: 1–22. http://doi:10.1029/2022GB007495

LaBuhn, S., and J. V. Klump. 2016. Estimating summertime epilimnetic primary production via in situ monitoring in an eutrophic freshwater embayment, Green Bay, Lake Michigan. J. Great Lakes Res. 42: 1026–1035. doi:10.1016/j.jglr.2016.07.028

Ladwig, R., A. P. Appling, A. Delany, H. A. Dugan, Q. Gao, N. Lottig, J. Stachelek, and P. C. Hanson. 2022. Long-term change in metabolism phenology in north temperate lakes. Limnol. Oceanogr. 67: 1502-1521. http://doi:10.1002/lno.12098

Lapierre, J. F., S. M. Collins, D. A. Seekell, and others. 2018. Similarity in spatial structure constrains ecosystem relationships: Building a macroscale understanding of lakes. Glob. Ecol. Biogeogr. 27: 1251–1263. http://doi:10.1111/geb.12781

Lapierre, J. F., D. A. Seekell, and P. A. del Giorgio. 2015. Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. Glob. Change Biol. 21: 4425–4435. http:// doi:10.1111/gcb.13031

Lapierre, J. F., and P. A. del Giorgio. 2014. Partial coupling and differential regulation of biologically and photochemically labile dissolved organic carbon across boreal aquatic networks. Biogeosciences 11: 5969–5985. http://doi:10.5194/bg-11-5969-2014

Lapierre, J. F., F. Guillemette, M. Berggren, and P. A. del Giorgio. 2013. Increases in terrestrially derived carbon stimulate organic carbon processing and CO2 emissions in boreal aquatic ecosystems. Nat. Commun. 4. http://doi:10.1038/ncomms3972

Lee, Z. P., K. P. Du, and R. Arnone. 2005. A model for the diffuse attenuation coefficient of downwelling irradiance. J. Geophys. Res. Oceans 110. http://doi:10.1029/2004JC002275

Liu, G., S. Li, K. Song, X. Wang, Z. Wen, T. Kutser, P. A. Jacinthe, Y. Shang, L. Lyu, C. Fang, and Y. Yang. 2021. Remote sensing of CDOM and DOC in alpine lakes across the Qinghai-Tibet Plateau using Sentinel-2A imagery data. J. Environ. Manage. 286: 112231. http://doi:10.1016/j.jenvman.2021.112231

Liu, S., and others. 2020. Linking the molecular composition of autochthonous dissolved organic matter to source identification for freshwater lake ecosystems by combination of optical spectroscopy and FT-ICR-MS analysis. Sci. Total Environ. 703. http://doi:10.1016/j.scitotenv.2019.134764

Liu, S., J. Hou, C. Suo, J. Chen, X. Liu, R. Fu, and F. Wu. 2022. Molecular-level composition of dissolved organic matter in distinct trophic states in Chinese lakes: Implications for eutrophic lake management and the global carbon cycle. Water Res. 217: 118438. http://doi:10.1016/j.watres.2022.118438

Lu, Y., Y. Gao, J. Jia, S. Wang, X. Ha, and Z. Li. 2023. Determining whether an 18/16O technology threshold exists in estimating lake primary productivity and associated metabolism. Ecol. Indic. 155: 110956. http://doi:10.1016/j.ecolind.2023.110956

MacCallum, R. C., and J. T. Austin. 2000. Applications of Structural Equation Modeling in Psychological Research. Annu. Rev. Psychol. 51: 201–226. http://doi:10.1146/annurev.psych.51.1.201

Mangal, V., Y. X. Shi, and C. Guéguen. 2017. Compositional changes and molecular transformations of dissolved organic matter during the arctic spring floods in the lower Churchill watershed (Northern Manitoba, Canada). Biogeochemistry 136: 151–165. http://doi:10.1007/s10533-017-0388-8

Mann, P. J., R. G. M. Spencer, P. J. Hernes, J. Six, G. R. Aiken, S. E. Tank, and others. 2016. Pan-arctic trends in terrestrial dissolved organic matter from optical measurements. Front. Earth Sci. 4: 1–18. http://doi:10.3389/feart.2016.00025

Marchand, D., Y. T. Prairie, and P. A. del Giorgio. 2009. Linking forest fires to lake metabolism and carbon dioxide emissions in the boreal region of Northern Québec. Glob. Chang. Biol. 15: 2861–2873. http://doi:10.1111/j.1365-2486.2009.01979.x

Massicotte, P., E. Asmala, C. Stedmon, and S. Markager. 2017. Global distribution of dissolved organic matter along the aquatic continuum: Across rivers, lakes and oceans. Sci. Total Environ. 609: 180–191. http://doi:10.1016/j.scitotenv.2017.07.076

Massicotte, P. 2016. eemR: Tools for Pre-Processing Emission-Excitation-Matrix (EEM) Fluorescence Data. Retrieved from https://cran.r-project.org/web/packages/eemR/

McDonald, C. P., E. G. Stets, R. G. Striegl, and D. Butman. 2013. Inorganic carbon loading as a primary driver of dissolved carbon dioxide concentrations in the lakes and reservoirs of the contiguous United States. Glob. Biogeochem. Cycles 27: 285–295. http://doi:10.1002/gbc.20032

McDonough, L. K., D. M. O'Carroll, K. Meredith, and others. 2020. Changes in groundwater dissolved organic matter character in a coastal sand aquifer due to rainfall recharge. Water Res. 169: 115201. http://http://doi: 10.1016/j.watres.2019.115201

Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. 7: 1–11. http://doi:10.1038/ncomms13603

Minor, E. C., M. M. Swenson, B. M. Mattson, and A. R. Oyler. 2014. Structural characterization of dissolved organic matter: A review of current techniques for isolation and analysis. Environ. Sci. Process. Impacts 16: 2064–2079. http://doi:10.1039/c4em00062e

Murphy, K. R., C. A. Stedmon, P. Wenig, and R. Bro. 2014. OpenFluor- An online spectral library of auto-fluorescence by organic compounds in the environment. Anal. Methods 6: 658–661. http://doi:10.1039/c3ay41935e

Murphy, K. R., C. A. Stedmon, D. Graeber, and R. Bro. 2013. Fluorescence spectroscopy and multi-way techniques. PARAFAC. Anal. Methods 5: 6557–6566. http://doi:10.1039/c3ay41160e

Nebbioso, A., and A. Piccolo. 2013. Molecular characterization of dissolved organic matter (DOM): A critical review. Anal. Bioanal. Chem. 405: 109–124. http://doi:10.1007/s00216-012-6363-2

Nürnberg, G. K., and M. Shaw. 1998. Productivity of clear and humic lakes: Nutrients, phytoplankton, bacteria. Hydrobiologia 382: 97–112. http://doi:10.1023/A:1003444827215

Ogdahl, M. E., V. L. Lougheed, R. J. Stevenson, and A. D. Steinman. 2010. Influences of multi-scale habitat on metabolism in a coastal Great Lakes watershed. Ecosystems 13: 222–238. http://doi:10.1007/s10021-009-9312-y

Ohno, T. 2002. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. Environ. Sci. Technol. 36: 742–746. http://doi:10.1021/es0155276

Oleksy, I. A., S. E. Jones, C. T. Solomon, and others. 2024. Controls on lake pelagic primary productivity: Formalizing the nutrient-color paradigm. J. Geophys. Res. Biogeosci. 129, e2024JG008140. http://doi:10.1029/2024JG008140

Oleksy, I. A., S. E. Jones, and C. T. Solomon. 2021. Hydrologic setting dictates the sensitivity of ecosystem metabolism to climate variability in lakes. Ecosystems. http://doi:10.1007/s10021-021-00718-5

Orlova, J., F. Amiri, A. K. Bourgeois, J. M. Buttle, E. Cherlet, C. W. Cuss, and others. 2024. Composition of Stream Dissolved Organic Matter Across Canadian Forested Ecozones Varies in Three Dimensions Linked to Landscape and Climate. Water Resour. Res. 60: 1–25. http://doi:10.1029/2023WR035196

Osburn, C. L., C. R. Wigdahl, S. C. Fritz, and J. E. Saros. 2011. Dissolved organic matter composition and photoreactivity in prairie lakes of the U.S. Great Plains. Limnol. Oceanogr. 56: 2371–2390. http://doi:10.4319/lo.2011.56.6.2371

Pace, M. L., and Y. T. Prairie. 2005. Respiration in lakes, p. 103–121. In P. A. del Giorgio and P. J. B. Williams [eds.], Respiration in aquatic ecosystems. Oxford Univ. Press. http://doi:10.1093/acprof:oso/9780198527084.003.0007

Paquette, C., I. Gregory-Eaves, and B. E. Beisner. 2021. Multi-scale biodiversity analyses identify the importance of continental watersheds in shaping lake zooplankton biogeography. J. Biogeogr. 48: 2298–2311. http://doi:10.1111/jbi.14153

Pugh, E. A., D. Olefeldt, S. N. Leader, K. J. Hokanson, and K. J. Devito. 2021. Characteristics of Dissolved Organic Carbon in Boreal Lakes: High Spatial and Inter-Annual Variability Controlled by Landscape Attributes and Wet-Dry Periods. Water Resour. Res. 57: 1–20. http://doi:10.1029/2021WR030021

Prairie, Y. T. 2008. Carbocentric limnology: Looking back, looking forward. Can. J. Fish. Aquat. Sci. 65: 543–548. http://doi:10.1139/F08-011

Prairie, Y. T., D. F. Bird, and J. J. Cole. 2002. The summer metabolic balance in the epilimnion of southeastern Quebec lakes. Limnol. Oceanogr. 47: 316–321. http://doi:10.4319/lo.2002.47.1.0316

Puts, I. C., J. Ask, A. Deininger, A. Jonsson, J. Karlsson, and A. K. Bergström. 2023. Browning affects pelagic productivity in northern lakes by surface water warming and carbon fertilization. Glob. Change Biol. 29: 375–390. http://doi:10.1111/gcb.16469

Puts, I. C., J. Ask, M. B. Siewert, R. A. Sponseller, D. O. Hessen, and A. K. Bergström. 2022. Landscape determinants of pelagic and benthic primary production in northern lakes. Glob. Chang. Biol. 28: 7063–7077. http://doi:10.1111/gcb.16409

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rivera Vasconcelos, F., S. Diehl, P. Rodríguez, J. Karlsson, and P. Byström. 2018. Effects of terrestrial organic matter on aquatic primary production as mediated by pelagic–benthic resource fluxes. Ecosystems 21: 1255–1268. http://doi:10.1007/s10021-017-0217-x

Rodríguez-Cardona, B. M., D. Houle, S. Couture, and others. 2023. Long-term trends in carbon and color signal uneven browning and terrestrialization of northern lakes. Commun. Earth Environ. 4, 338. http://doi:10.1038/s43247-023-00999-9

Rosseel, Y. 2012. Lavaan: An R package for structural equation modeling. J. Stat. Softw. 48. http://doi:10.18637/jss.v048.i02

Ryan, K. A., V. A. Garayburu-Caruso, B. C. Crump, T. Bambakidis, P. A. Raymond, S. Liu, and J. C. Stegen. 2024. Riverine dissolved organic matter transformations increase with watershed area, water residence time, and Damköhler numbers in nested watersheds. Biogeochemistry 167: 1203–1224. http://doi:10.1007/s10533-024-01169-5

Sadro, S., and J. M. Melack. 2012. The effect of an extreme rain event on the biogeochemistry and ecosystem metabolism of an oligotrophic high-elevation lake. Arct. Antarct. Alp. Res. 44: 222–231. http://doi:10.1657/1938-4246-44.2.222

Saros, J. E., V. Huston, R. M. Northington, G. P. Lamb, A. Birkel, S. Pienkowski, G. Bourdin, B. Jiang, and S. Michaud. 2025. Abrupt transformation of west Greenland lakes following compound climate extremes associated with atmospheric rivers. Proc. Natl. Acad. Sci. U.S.A. 122, e2023456789. http://doi:10.1073/pnas.2023456789

Sargent, M. C., and T. S. Austin. 1949. Organic productivity of an Atoll. Eos Trans. Am. Geophys. Union 30: 245–249. http://doi:10.1029/TR030i002p00245

Schindler, D. W., S. E. Bayley, B. R. Parker, and others. 1996. The effects of climatic warming on the properties of boreal lakes and streams at the Experimental Lakes Area, northwestern Ontario. Limnol. Oceanogr. 41: 1004-1017. http://doi:10.4319/lo.1996.41.5.1004

Scordo, F., N. R. Lottig, J. E. Fiorenza, and others. 2022. Hydroclimate variability affects habitat-specific (open water and littoral) lake metabolism. Water Resour. Res. 58. http://doi:10.1029/2021WR031094

Seekell, D. A., J. F. Lapierre, and K. S. Cheruvelil. 2018. A geography of lake carbon cycling. Limnol. Oceanogr. Lett. 3: 49–56. http://doi:10.1002/lol2.10078

Seekell, D. A., J. F. Lapierre, J. Ask, A. Bergstreom, A. Deininger, P. Rodriguez, and J. Karlsson. 2015. The influence of dissolved organic carbon on primary production in northern lakes. Limnol. Oceanogr. http://doi:10.1002/lno.10096

Seymour, J. R., S. A. Amin, J. B. Raina, and R. Stocker. 2017. Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. Nat. Microbiol. 2. http://doi:10.1038/nmicrobiol.2017.65

Smith, E. M., and W. M. Kemp. 1995. Seasonal and regional variations in plankton community production and respiration for Chesapeake Bay. Mar. Ecol. Prog. Ser. 119: 217–231. http://doi:10.3354/meps119217

Smith, H. J., M. Tigges, J. D'Andrilli, A. Parker, B. Bothner, and C. M. Foreman. 2018. Dynamic processing of DOM: Insight from exometabolomics, fluorescence spectroscopy, and mass spectrometry. Limnol. Oceanogr. Lett. 3: 225–235. http://doi:10.1002/lol2.10082

Smith, S. V., J. T. Hollibaugh, S. J. Dollar, and S. Vink. 1991. Tomales Bay metabolism: C-N-P stoichiometry and ecosystem heterotrophy at the land-sea interface. Estuar. Coast. Shelf Sci. 33: 223–257. http://doi:10.1016/0272-7714(91)90055-G

Smits, A. P., F. Scordo, M. Tang, and others. 2024. Wildfire smoke reduces lake ecosystem metabolic rates unequally across a trophic gradient. Commun. Earth Environ. http://doi:10.1038/s43247-024-01404-9

Smol, J. P., A. P. Wolfe, H. J. B. Birks, and others. 2005. Climate-driven regime shifts in the biological communities of arctic lakes. Proc. Natl. Acad. Sci. U.S.A. 102: 4397-4402. http://doi:10.1073/pnas.0500245102

Sobek, S., L. J. Tranvik, Y. T. Prairie, P. Kortelainen, and J. J. Cole. 2007. Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. Limnol. Oceanogr. 52: 1208–1219. http://doi:10.4319/lo.2007.52.3.1208

Solomon, C. T., S. E. Jones, B. C. Weidel, and others. 2015. Ecosystem consequences of changing inputs of terrestrial dissolved organic matter to lakes: Current knowledge and future challenges. Ecosystems 18: 376–389. http://doi:10.1007/s10021-015-9848-y

Solomon, C. T., D. A. Bruesewitz, D. C. Richardson, and others. 2013. Ecosystem respiration: Drivers of daily variability and background respiration in lakes around the globe. Limnol. Oceanogr. 58: 849–866. http://doi:10.4319/lo.2013.58.3.0849

Søndergaard, M., C. A. Stedmon, and N. H. Borch. 2003. Fate of terrigenous dissolved organic matter (DOM) in estuaries: Aggregation and bioavailability. Ophelia 57: 161–176. http://doi:10.1080/00785236.2003.10409512

Soranno, P. A., K. S. Cheruvelil, E. G. Bissell, and others. 2014. Cross-scale interactions: Quantifying multi-scaled cause-effect relationships in macrosystems. Front. Ecol. Environ. 12: 65–73. http://doi:10.1890/120366

Sothe, C., A. Gonsamo, J. Arabian, W. A. Kurz, S. A. Finkelstein, and J. Snider. 2022. Large soil carbon storage in terrestrial ecosystems of Canada. Glob. Biogeochem. Cycles 36. http://doi:10.1029/2021GB007213

Spencer, R. G. M., A. M. Kellerman, D. C. Podgorski, and others. 2019. Identifying the molecular signatures of agricultural expansion in Amazonian headwater streams. J. Geophys. Res. Biogeosci. 124: 1637–1650. http://doi:10.1029/2018JG004910

Spencer, R. G. M., K. D. Butler, and G. R. Aiken. 2012. Dissolved organic carbon and chromophoric dissolved organic matter properties of rivers in the USA. J. Geophys. Res. Biogeosci. 117. http://doi:10.1029/2011JG001928

Stadler, M., E. Ejarque, and M. J. Kainz. 2020. In-lake transformations of dissolved organic matter composition in a subalpine lake do not change its biodegradability. Limnol. Oceanogr. 1–19. http://doi:10.1002/lno.11406

Staehr, P. A., L. Baastrup-Spohr, K. Sand-Jensen, and others. 2012. Lake metabolism scales with lake morphometry and catchment conditions. Aquat. Sci. 74: 155-169. http://doi:10.1007/s00027-011-0207-6

Staehr, P. A., K. Sand-Jensen, A. L. Raun, B. Nilsson, and J. Kidmose. 2010. Drivers of metabolism and net heterotrophy in contrasting lakes. Limnol. Oceanogr. 55: 817–830. http://doi:10.4319/lo.2009.55.2.0817

Staehr, P. A., and K. Sand-Jensen. 2007. Temporal dynamics and regulation of lake metabolism. Limnol. Oceanogr. 52: 108–120. http://doi:10.4319/lo.2007.52.1.0108

Stedmon, C. A., R. M. W. Amon, A. J. Rinehart, and S. A. Walker. 2011. The supply and characteristics of colored dissolved organic matter (CDOM) in the Arctic Ocean: Pan Arctic trends and differences. Mar. Chem. 124: 108–118. http://doi:10.1016/j.marchem.2010.12.007

Stedmon, C. A., and R. Bro. 2008. Characterizing dissolved organic matter fluorescence with PARAFAC: a tutorial. Limnol. Oceanogr. Methods 6: 572–579. http://doi:10.4319/lom.2008.6.572

Stedmon, C. A., S. Markager, and R. Bro. 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Mar. Chem. 82: 239–254. http://doi:10.1016/S0304-4203(03)00072-0

Stekhoven, D. J., and P. Bühlmann. 2012. MissForest—Non-parametric missing value imputation for mixed-type data. Bioinformatics 28: 112–118. http://doi:10.1093/bioinformatics/btr597

Stolpmann, L., C. Coch, A. Morgenstern, J. Boike, M. Fritz, U. Herzschuh, and others. 2021. First pan-Arctic assessment of dissolved organic carbon in lakes of the permafrost region. Biogeosciences 18: 3917–3936. http://doi:10.5194/bg-18-3917-2021

Stubbins, A., J. F. Lapierre, M. Berggren, Y. T. Prairie, T. Dittmar, and P. A. del Giorgio. 2014. What's in an EEM? Molecular signatures associated with dissolved organic fluorescence in boreal Canada. Environ. Sci. Technol. 48: 10598–10606. http://doi:10.1021/es502086e

Stubbins, A., V. Hubbard, G. Uher, and others. 2008. Relating carbon monoxide photoproduction to dissolved organic matter functionality. Environ. Sci. Technol. 42: 3271–3276. http://doi:10.1021/es703014q

Tanentzap, A. J., A. Fitch, C. Orland, E. J. S. Emilson, K. M. Yakimovich, H. Osterholz, and T. Dittmar. 2019. Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. Proc. Natl. Acad. Sci. U.S.A. 116: 24689–24695. http://doi:10.1073/pnas.1904896116

Tanentzap, A. J., and others. 2017. Terrestrial support of lake food webs: Synthesis reveals controls over cross-ecosystem resource use. Sci. Adv. 3. http://doi:10.1126/sciadv.1601765

Turner, K. W., T. W. D. Edwards, and B. B. Wolfe. 2014. Characterizing runoff generation processes in a lake-rich thermokarst landscape (Old Crow Flats, Yukon, Canada) using δ18O, δ2H and d-excess measurements. Permafrost Periglac. Process. 25: 53–59. http://doi:10.1002/ppp.1802

U.S. Environmental Protection Agency (USEPA). 2009. National Lakes Assessment: A collaborative survey of the nation's lakes. EPA-841-R-09-001. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC.

Vachon, D., S. Sadro, M. J. Bogard, and others. 2020. Paired O2–CO2 measurements provide emergent insights into aquatic ecosystem function. Limnol. Oceanogr. Lett. 5: 287–294. http://doi:10.1002/lol2.10135

Vanderwall, J. W., A. Ballantyne, and J. J. Elser. 2024. Net ecosystem metabolism is independent of elevation in mountain lakes of the northern Rocky Mountains, USA. Ecosystems 27: 1040–1059. http://doi:10.1007/s10021-024-00935-8

Verpoorter, C., T. Kutser, D. A. Seekell, and L. J. Tranvik. 2014. A global inventory of lakes based on high-resolution satellite imagery. Geophys. Res. Lett. 41: 6396–6402. http://doi:10.1002/2014GL060641

Wagner, K., M. M. Bengtsson, R. H. Findlay, T. J. Battin, and A. J. Ulseth. 2017. High light intensity mediates a shift from allochthonous to autochthonous carbon use in phototrophic stream biofilms. J. Geophys. Res. Biogeosci. 122: 1806–1820. http://doi:10.1002/2016JG003727

Wang, T., W. Feng, J. Liu, W. Fan, T. Li, F. Song, and others. 2024. Eutrophication in cold-arid lakes: molecular characteristics and transformation mechanism of DOM under microbial action at the ice-water interface. Carbon Res. 3. http://doi:10.1007/s44246-024-00126-z

Ward, N. D., T. S. Bianchi, P. M. Medeiros, M. Seidel, J. E. Richey, R. G. Keil, and H. O. Sawakuchi. 2017. Where Carbon Goes When Water Flows: Carbon Cycling across the Aquatic Continuum. Front. Mar. Sci. 4: 1–27. http://doi:10.3389/fmars.2017.00007

Wen, Z., Y. Shang, K. Song, G. Liu, J. Hou, L. Lyu, and others. 2022. Composition of dissolved organic matter (DOM) in lakes responds to the trophic state and phytoplankton community succession. Water Res. 224: 119073. http://doi:10.1016/j.watres.2022.119073

Weyhenmeyer, G. A., and D. J. Conley. 2017. Large differences between carbon and nutrient loss rates along the land to ocean aquatic continuum—implications for energy:nutrient ratios at downstream sites. Limnol. Oceanogr. 62: S183–S193. http://doi:10.1002/lno.10589

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee, G. L. Smith, and J. E. Cooper. 1996. Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. Bull. Am. Meteorol. Soc. 77: 853–868. http://doi:10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2

Williamson, C. E., D. P. Morris, M. L. Pace, and O. G. Olson. 1999. Dissolved organic carbon and nutrients as regulators of lake ecosystems: Resurrection of a more integrated paradigm. Limnol. Oceanogr. 44: 795–803. http://doi:10.4319/lo.1999.44.3_part_2.0795

Winterdahl, M., K. Bishop, and M. Erlandsson. 2014. Acidification, dissolved organic carbon (DOC) and climate change, p. 1–7. In B. Freedman [ed.], Global environmental change. Springer. http://doi:10.1007/978-94-007-5784-4_107

Woolway, R. I., S. Sharma, and J. P. Smol. 2022. Lakes in hot water: The impacts of a changing climate on aquatic ecosystems. BioScience 72: 1050–1061. http://doi:10.1093/biosci/biac052

Xenopoulos, M. A., R. T. Barnes, K. S. Boodoo, D. Butman, N. Catalán, S. C. D'Amario, and others. 2021. How humans alter dissolved organic matter composition in freshwater: relevance for the Earth's biogeochemistry. Biogeochemistry 154: 323–348. http://doi:10.1007/s10533-021-00753-3

Yan, M., G. Korshin, D. Wang, and Z. Cai. 2012. Characterization of dissolved organic matter using high-performance liquid chromatography (HPLC)-size exclusion chromatography (SEC) with a multiple wavelength absorbance detector. Chemosphere 87: 879–885. http://doi:10.1016/j.chemosphere.2012.01.029

Yvon-Durocher, G., J. M. Caffrey, A. Cescatti, and others. 2012. Reconciling the temperature dependence of respiration across timescales and ecosystem types. Nature 487: 472–476. http://doi:10.1038/nature11205

Yvon-Durocher, G., J. I. Jones, M. Trimmer, G. Woodward, and J. M. Montoya. 2010. Warming alters the metabolic balance of ecosystems. Philos. Trans. R. Soc. B 365: 2117–2126. http://doi:10.1098/rstb.2010.0038

Zwart, J. A., and L. S. Brighenti. 2021. Measurement and variability of lake metabolism. In Encyclopedia of Inland Waters (Second Edition), 2: 163–173. http://doi:10.1016/B978-0-12-819166-8.00029-3

Zwart, J. A., N. Craig, P. T. Kelly, S. D. Sebestyen, C. T. Solomon, B. C. Weidel, and S. E. Jones. 2016. Metabolic and physiochemical responses to a whole-lake experimental increase in dissolved organic carbon in a north-temperate lake. Limnol. Oceanogr. 61: 723–734. http://doi:10.1002/lno.10248