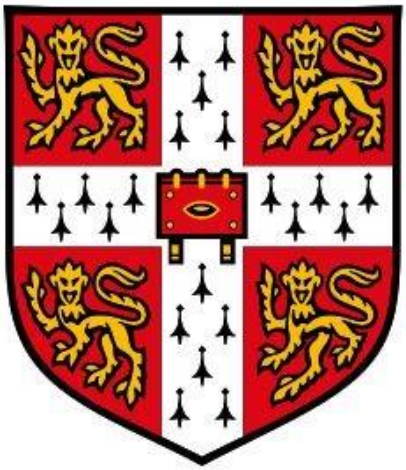
Logging impacts the ecology of molecules in headwater streams



Erika Christiane Freeman

Lucy Cavendish College

University of Cambridge

January 2023

This dissertation is submitted for the degree of Doctor of Philosophy

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the 60,000 Degree Committee

For those close to my heart who have patiently stood by while I was off figuring out how to be a scientist

and who have humbled me by summarising my years of work better than I ever could: “something about dirt water.”

Every rill is a channel for the juices of the meadow. Last year's grasses and flower-stalks have been steeped in rain and snow, and now the brook flows with meadow tea—thoroughwort, mint, flagroot, and pennyroyal, all at one draught.

Henry David Thoreau, March 8, 1840

Acknowledgments

Coming to the University of Cambridge as a first-generation graduate student felt like I had been ushered into a dream, especially the life-changing magic of being included amongst the Gates Cambridge Scholars. I suddenly found myself surrounded by some of the most talented and kind people I have ever met. Few could predict how the world would change as the threat of COVID emerged. I spent much of the pandemic on the move, finding my way through Canadian streams, bogs, and forests, to the channels and oaks in Oldenburg, the intricacies of the sub-visible, the opulent and pristine streets of Zürich, and back again. I learned so much about the natural world, but I probably learned the most about myself—these lessons I will carry with me for the rest of my life.

This experience would not have been possible without my advisors, Andrew and Erik.

Andrew, I came to Cambridge with little more than a hunger to learn, half-baked ideas, and the resolve to work hard. I am incredibly fortunate to have had a chance to glimpse into the inner workings of a truly great scientist. Your attention to detail, commitment to excellence, and draft feedback (rate and precision) have been otherworldly. Anytime I have given you any written work, of which there was a wide range of qualities, I felt as if I arrived with a block of lumpy and misshaped clay, and you help me mould it into something recognisable, if not beautiful. I am also in awe of your strength and leadership amidst a global crisis. Even before the pandemic arrived, I had not met someone who so expertly avoided contact with doorknobs or the gentle murmurs of an in-coming cold. You were ahead of the curve, and I have admired your foresight and good judgement and probably always will. Above all, thank you for an incredible amount of support during this difficult time. I will always be grateful for the effort you put into helping me see this through.

Erik, I showed up on your doorstep with little more than an insanely ambitious project plan and the resolve to work hard. You navigated me along a steep learning curve and buoyed me through many weeks and months of gruelling field and lab work. I often still laugh at the norm-bending amount of work we managed to pull off: sampling 200 streams, setting up experimental catchments with hundreds of lysimeters and then

creating a harvest experiment from scratch, building an isotope lab (also from scratch), finding and fixing a hydrogen isotope analyser, digging ditches in the forest for metagenomics data. There are too many even to list. But, most of all, for your support, good humour, and kindness as I tried to find my way. The moment you told me that Elodie was to arrive was one of my most cherished memories from 2020, I hope she grows up to be like her parents.

I hope to continue learning from both of you into the future and look forward to seeing you again someday on home soil.

This Ph.D. would never have been possible without the financial support of my generous sponsors, especially the Gates Cambridge Trust. I am also exceptionally grateful for the financial and administrative support of the Department of Plant Sciences throughout my time there. Lastly, I thank my college, Lucy Cavendish, which provided financial support in the form of travel grants and assigned me the best tutor I could have hoped for. Thank you, Howard Nelson, for your encouragement and kindness.

I also want to thank the many people who helped me on my quixotic quest. Thanks to Thorsten Dittmar for the engaging conversations and incredible generosity in letting me spend months in your lab processing samples. To Katrin and Ina, for your friendship, chats, and support. You are two of the kindest people I have encountered in all of my travels. Bis bald, hoffentlich.

Thank you, Irena Creed, for sending me field support and for the countless ways you have supported and bolstered my career. To the good folks at the Canadian Forest Service (NRCan-CFS) for their considerable support. Especially: Caroline Emilson, Paul Hazlet, Kara Webster, Jason Leach, Emily Smenderovac, Derek Chartrand, Brian Keilstra, Kristi Broad, Kerrie Wainio Keizer, Linda Vogel, Laura Hawdon, Ken McIlwrick, Scott Capell, and so many more. To the entire engineering team for entertaining my requests and building the best Ferrari cooler the world has ever seen. Without exaggeration, it zoomed along with me at outrageous speeds on the Autobahn to Oldenburg. Thank you to the security team for turning the other way when I was at work far later than we were supposed to be. Thank you to the custodial staff for the incredible rapport. I always knew you had my back. A special thank you is to Joe Schadenberg for being the yin to my fieldwork yang. An additional special thanks to

Lucy, Kaia, and Will for one of the most memorable summers of my life. I owe you some peppers.

To members of the ecology group, we did not spend much time together in the same place, but it was a treasured time! A special thanks to Chloe Orland, Sasha Khoury, Danai Kontou, Sarah Sandor, Laura Bently, Jeremy Fonvielle, Javier Igea, Jon Williams, Tom Swinfield, Yi Zhang, James Ball, David Coomes, and all other members of the Ecology groups with whom I have had the pleasure of crossing paths. Sam Woodman, a special thanks to you for all those Zoom calls to commiserate.

This was an unusually isolating time to complete a Ph.D., especially one that saw me move to 11 different homes, 4 different countries, many weeks in quarantine, and lonely journeys through the Canadian woods. I want to thank the people that saw me through those challenging times:

To my fellow Gates Scholars and friends, Kim Van Daalen, Chris van Hoorn, Vineet Nair, and June Park. Those first two years with you brought such value to my life. I can’t wait to see all the great things you get up to. A special thank you to my friends Ayan Mandal and Ethan Dutcher. While our skills in predicting primaries may be lacking, I am certain made the right decision about with whom to spend my time at Cambridge. Thank you for challenging me in the best possible ways. Thank you, Anne Thomas, you could have been featured many times, but I think it's most fitting here with close friends. Thank you for being my pen pal, confidant, and partner in fries during these exhilarating times. Thank you to the lovely people at Honest Burger for providing the (often free) fries.

To the other cast of characters who have brightened my life along the way: to Aaron for infusing my 2020 summer with magic and reminding me why I loved science. I hope you make your mark on this world, or perhaps even beyond it. Matt for the saving grace friendship. Daniel with a warm heart and playful attitude, meeting you was a decisive moment in my life. I will always be grateful. Adrian, it is unfair that there should be someone so simultaneously warm, athletic, fun, and competent. I am looking forward to more outrageous adventures. To Phil the best sola-duo half a person could ask for. To Evgeny, for the playful spirit and patience as I stomped all over your feet. Hopefully, there will be more time to dance. To Amy and Sam, for a summer filled with heart-to-hearts and most excellent adventures. Thank you to Ellis for your

companionship, a steady hand, and the countless drafts and rushed writing you sifted through. You brought me so much joy. Katrina, your sweet emails, editorial prowess, and, above all, bringing Ellis into this world, have my deepest appreciation. To Ana, Ceci, Inez, and Leti. Thank you for the beautiful people that you are inside and out. Your support and time have meant so much to me and I could not have done it without you. To my oldest friends (by length but also maybe by age too): Michelle, Teagan, and Riley. For reminding me who I was before I tried to do a Ph.D.

To my grandparents for your unyielding support and encouragement in my academic pursuits. My uncle, Sigi, for your steadfast unterstützung and patience. To Sandra, for your sweet messages and encouragement during my most indecisive and uncertain times. To my younger siblings: Karin, Patrick, and Jonny. I love you more than anything. To my Dad, for being there when I needed you, for the good humour, and for bailing me out of Swiss medical bankruptcy. Finally, thank you to my nervous system. I am so sorry for what I did to you.

This journey has been filled with countless challenges, but it has also been a joyous one thanks to the incredible learning and people who have touched my life along the way. I am forever grateful for the support, love, and kindness I have received. The memories and lessons from this experience will stay with me for the rest of my life. Lastly to you, dear reader. I hope you learn something new.

Summary

Forest management and harvest offer a promising means of combating climate change by removing CO_2 from the atmosphere. However, most forest carbon (C) is held in soils. Thus, by disturbing soils and altering hydrology, forest management and harvest potentially displace large amounts of C from forest soils into aquatic ecosystems. My dissertation seeks to understand the fate of this forest C as dissolved organic matter (DOM) into various aquatic endpoints by tracing its molecular composition along the ephemeral water film that begins in upland soils and ends in streams. The fate and function of DOM in aquatic ecosystems are strongly affected by its chemical properties. Thus, recent progress in the molecular characterization of DOM has opened a new line of inquiry into harvest impacts on aquatic ecosystem functioning. My thesis advances this line of inquiry by applying high-resolution mass spectrometry to study the effect of forest disturbance on DOM in soils and connected streams.

Chapter 1 of this thesis gives a general overview of the molecular nature of terrestrial DOM sources, how these sources may be altered by harvest, and the subsequent transfer, fate, and properties of DOM once in streams. It also outlines the specific objectives I address with a combination of field experiments and synoptic surveys in the Batchewana watershed in Ontario, Canada. In **Chapter 2,** I tracked DOM along soil depths, and hillslope positions in four replicate forest headwater catchments of the Canadian hardwood forest. I related DOM composition to soil microbiomes and physical chemistry to establish baseline conditions before a harvest experiment. I found that DOM changed similarly along soil-aquatic gradients, irrespective of differences in environmental conditions. My results implicated continuous microbial reworking that shifts DOM towards a shared pool of compounds in soils. Such general degradation patterns can inform the management of soil-to-stream carbon losses by predicting DOM composition and its downstream reactivity along environmental gradients. In **Chapter 3**, I quantified and characterised the effects of logging on DOM composition over three years using the four experimental catchments from **Chapter 2**. Two catchments were experimentally logged, while the remaining two were left as controls. I found that DOM concentration in stream water

from logged catchments increased in a pulse during the first year, but only the changes in the quality of DOM persisted. Using ultrahigh-resolution mass spectrometry, I showed that DOM released from deforested catchments was energy-rich and more chemically diverse, likely because of higher hydrological connectivity with intermediate and deep soil layers. I estimated that while logging increased the overall annual flux of dissolved organic carbon by approximately 8.5% of the extracted wood carbon, the exposure of deeper soil through logging released previously stable soil organic carbon to streams. The resulting changes to the molecular composition of DOM within headwater streams persisted for at least two years after logging, potentially disrupting aquatic ecosystems and making streams more likely to release terrestrial C into the atmosphere. In **Chapter 4**, I examined the chemical properties of DOM in stream water from over 200 Canadian headwater streams in an area with historical forest harvest. I demonstrated that using the fluorescence properties of streams, the effect of harvest, although detectable on a large spatial scale, is relatively minor compared to the effects of forest types and wetness gradients. These results have implications for land-water linkages under a changing climate that shift terrestrial sources of DOM. Finally, in **Chapter 5**, I discuss the implications of my findings for understanding the coupling between terrestrial and aquatic ecosystems and propose avenues for future research.

Table of Contents

Table of Figures

Table of Tables

Appendices

Preface

To facilitate the publication of the thesis, **Chapters 2-4** are written as manuscripts for peer reviewed journals. Notably, these chapter are written with the Materials and Methods section located after the Results and Discussion. Contributions from co-authors include field assistance, laboratory support, data provisions, and supervisory guidance. Therefore, I use the pronoun “we” rather than “I” when referring to collective work throughout the thesis. Author contributions for each chapter are as follows:

Chapter 1 is entirely my own work.

Chapter 2 is currently out for review at *Nature Communications*: **Freeman, EC**, Emilson, EJS, Braga, L, Emilson, C, Martineau, C, Dittmar, T, Goldhammer, T, Singer, G., Tanentzap AJ. (2023). Universal Microbial Processes Rework Dissolved Organic Matter along Environmental Gradients. EJSE, AJT. and I conceived of the study. I performed fieldwork, collected samples, and set-up field sites with input from EJS and AJT and performed FT-ICR-MS measurements assisted by TD. TG and SG processed LC-OCD data. LB, CE, and CM processed genomics samples and performed bioinformatics. I processed and analyzed all data with input from AJT. I wrote the manuscript with input from AJT and all authors.

Chapter 3 is out for review at *Nature Geoscience*: **Freeman, EC,** Emilson, JSE, Webster, K., Dittmar, T., Tanentzap, A. (2023). Forest Harvest Disrupts the Ecology of Molecules in Headwater Streams. EJSE, AJT and I conceived of the experiment. I established the field sites and collected samples guided by EJSE and AJT. I processed and analysed all FT-ICR-MS data assisted by TD. EJSE processed all optical data which I analysed with input from AJT. I wrote the manuscript with input from AJT and all authors.

Chapter 4 is currently in preparation for *Global Change Biology* and was conceived by AJT, EJSE and I. I collected all stream samples with input from EJSE. EJSE provided additional samples and performed PARAFAC modeling. I analysed and collected all samples for FT-ICR-MS and performed all geospatial analysis and processed and analysed all data with input from EJSE. I wrote the manuscript with input from EJSE and all authors.

Chapter 5 is entirely my own work.

In addition, the following papers have resulted from work conducted during this thesis are not included in the text:

Borton, MA, Collins, SM, Graham, EB, Garayburu-Caruso, VA, Goldman, AE, de Melo, M,... & **WHONDRS Crowdsourced Consortium**. It Takes a Village: Using a Crowdsourced Approach to Investigate Organic Matter Composition in Global Rivers Through the Lens of Ecological Theory (2022). Frontiers in Water.

Lembrechts, J, Van den Hoogen, J, Aalto, J, Ashcroft, M, De Frenne, P, Kemppinen J, **Freeman, EC**,... & Hoffrén R. Global Maps of Soil Temperature (2021). Global Change Biology

Buser-Young, J, Garcia, P, Schrenk, M, Regier, P, Ward, ND, Biçe, K, Brooks, S, **Freeman, EC**, Lønborg, C. Determining the Biogeochemical Transformations of Organic Matter Composition in Global Rivers Using Molecular Signatures (2022). Frontiers in Water

Stadler, M, de Melo, M, Garayburu-Caruso, V, **Freeman, EC**, Bice, K, Mateus-Barros, E, Lu, Y, Barnard, M, Cheng, S, Dwivedi, D, Tanentzap, AJ, Meile, C. Exploring Global River Corridor OM Chemistry through the Ecological Concept of Core-satellite Species (2023).
Frontiers in Water

Freeman EC, Mudunuru ML , Ward CS, Mudunuru ML, Bottos EM, McClure EA, González-Pinzón R, Feeser KL, Krause S, Peña J, Newcomer ME. A Unified Conceptual Model of Organic Matter Scaling in River Corridors.
In Preparation for Frontiers in Water

Chapter 1

General introduction

1.1 Forests as nature-based climate solutions and the increasing role of harvest

Forests are intrinsically linked to water, with forested watersheds providing 75 percent of our accessible freshwater resources[1,2]. Forests are natural reservoirs and filters that can store, release, and purify water through their interactions with hydrological and sediment transport processes and nutrient cycling at multiple scales[3–7]. Yet, globally, forests are in flux due to human activities (such as logging, reforestation, afforestation, agriculture, and urbanisation) and natural disturbances (insect outbreaks and wildfires). Terrestrial and freshwater research and watershed management must pay close attention to the relationships between forests and water. This is needed to protect the water-related benefits that forests provide in the face of environmental changes.

Furthermore, forests are also extremely critical in meeting the less than 2°C global warming goal outlined in the Paris Agreement by facilitating the removal of CO_2 from the atmosphere[8,9]. For example, forest carbon (C) can be sequestered by producing wood-derived building materials[10] and biochar[11]. These strategies work by storing C in long-lasting wood products; meanwhile, logged land can be used to regrow forests and draw down more CO_2[12] from the atmosphere.

However, we do not know the effectiveness of forest harvest and re-growth as a C sequestration method[13–15] or its trade-offs with other earth system components, such as inland waters. The potential ineffectiveness of this strategy is clear when considering that soil organic matter (OM) stores more C than woody vegetation[16]. Likewise, global soils store at least three times as much C as the atmosphere or plants[16]. Thus, if soil is increasingly disturbed by mechanised harvest as tree crops are turned over[17] then C loss to inland waters is expected due to erosion from

reduced surface cover and soil compaction[18]. This loss can double the amount of C in streams[19], though it is unknown what proportion of forest C sequestration this loss might undo. Once in water, soil C is much more likely to be returned to the atmosphere[20,21]. The flux of C from land to inland waters has been estimated to comprise anywhere between 12-34% of terrestrial net ecosystem production (without disturbance)[22,23], while the C returning to the atmosphere from inland waters is about 76% of the amount received from land[23]. Given that the Paris commitments rely, in significant part, on negative emissions and that afforestation and reforestation are the only mitigation measures that have been proven to be implementable at large scale, the calculus of earth system trade-offs must be examined carefully[24].

This dissertation aims to build the foundation for a risk assessment of forest biomass as a negative-emission technology[25]. In particular, this dissertation looks at the trade-offs between forest harvest for C sequestration and the potential water quality loss and C emissions from water (the forest-water nexus) through the lens of dissolved organic matter (DOM).

1.2 Loss of carbon from forests to waterways as DOM

In the previous section, I discussed "forests as a natural climate solution." In this section, I introduce and describe the notion of DOM, which, as I discuss in later sections, is highly affected by the harvest of forests as a natural climate solution.

DOM is commonly defined as soluble and colloidal organic material able to pass through a 0.45 μm filter[26] and is one of the most complex natural mixtures known[27,28]. It is a useful lens for exploring the forest-water nexus for four primary reasons relating to the source, size, and reactivity of the forest-water DOM flux. First, most (up to 90%) of the DOM in surface waters is derived from terrestrial vegetation and soils[29]. Therefore, understanding terrestrial DOM is the starting point to understanding most DOM in surface waters. Second, DOM accounts for the vast majority of organic C that flows through aquatic networks (rather than larger particulate OM)[20]. Therefore, DOM is the relevant C form if manipulations are to be developed to sequester C in waters. Third, DOM in inland waters has a relatively short mean half-life (4.7 ± s.d. 2.5 yr^{-1}) compared to its centennial residence times in soils[30]. Therefore, inland waters are increasingly recognized as efficient bioreactors

for terrestrial OM[31], and are more likely to release C to the atmosphere as CO_2 than soil. Finally, DOM comprises different classes of organic compounds with differing reactivities and ecological roles. Thus, if even a small proportion of this significantly sized and easily transformed flux is elevated or shifted to a more reactive form, it could slow our ability to meet CO_2 removal targets. Together, this raises the following questions:

How do our actions on land affect the DOM exported to waters? And subsequently, how do the features of terrestrial DOM determine its likelihood of being transformed and released to the atmosphere once in aquatic ecosystems?

1.3 The nature of terrestrial DOM

Core to answering these questions is an understanding of what terrestrial DOM is. Organic matter is stored on land within terrestrial biota (570 gigatons C; 1 Gt=1×1015 g), plant litter (70 Gt C[32]), and surface (0-100cm) soils (1600 Gt C[33]). The main source of terrestrial DOM is vascular plants, which are, in almost all cases, confined to land and contain characteristically high concentrations of nitrogen-free biomacromolecules such as lignin, tannin, suberin, and cutin[34]. Assuming that litter and biomass are compositionally similar[35,36], the total inventory of above-ground terrestrial plant matter corresponds roughly to 250 Gt C in cellulose, 175 Gt lignin C, 150 Gt of C in hemicellulose and other polysaccharides, and 5–10 Gt C each in protein, lipid, and cutin. Thus, the possible contribution of plant material is relatively straightforward (compared to the contribution of soils). However, there is a great deal of variation in terrestrial DOM among plant species[37] and ecological regions[38], explained more accurately and extensively with the rapidly expanding field of environmental metabolomics[39]. Thus these numbers presented must be considered a simplification.

Until recently, it was presumed that most soil organic matter (SOM) was derived from plant litter, i.e., the soil is built from the top down[40]. However, this view has evolved due to work demonstrating that root-derived C and C exuded by fungal communities associated with root systems can contribute up to 70% of the C in soils[41,42]. Thus, the main driver of SOM formation may not be leaf decomposition *per se* but microbial and fungal-derived C resulting from released metabolites and

necromass[43]. These findings imply that the properties of SOM will change non-linearly with depth and that relatively new and potentially labile compounds could exist throughout the soil column[44]. Therefore, if hotspots or larger areas of SOM become hydrologically connected, more labile DOM may end up in streams. For example, Kaiser et al. (2017) found that in large Arctic rivers, terrestrially-derived bacterial OM accounted for 21–42% of DOM in all watersheds[45].

Changes in SOM quantity and quality are mainly determined by three factors: types of organic input, biological activity, and environmental and edaphic variables[46]. The rates of primary production and the chemodiversity of plants determine the amount and type of organic carbon delivered into soils through litterfall, root turnover and exudation. This plant material then passes through the microbial “funnel” during the decomposition process[43] and is either released back to the atmosphere as CO_2, modified ex-vivo (extracellular) or accumulated as microbial by-products or necromass[43]. Furthermore, because plant inputs decrease with depth, SOM is increasingly decomposed as it passes through the soil profile[40,47,48]. In deeper mineral soils, physical and chemical processes, such as adsorption to mineral surfaces and protection from degradation, play a relatively more significant role compared to shallow organic horizons. Although microorganisms affect the accumulation and loss of organic matter in soils through many pathways, microbial carbon use efficiency (CUE), i.e., the microbial partitioning of carbon used in metabolisms that goes towards growth or respiration, has been shown to have a positive correlation with soil organic carbon content. CUE has been show to be at least four times as important as factors such as: plant carbon inputs, carbon input allocation, non-microbial carbon transfer (organo-mineral interactions), substrate decomposability, environmental modifications and vertical transport[49]. The relationship between CUE and SOM however dampens with soil depth, indicating stronger interactions of organic matter with mineral particles at depth[48].

Aquatic research often does not distinguish between soil and plant-derived DOM. Instead, terrestrial DOM was historically considered a monolith mainly composed of old, high molecular weight aromatic compounds (mostly humics)[50]. Once in water, this material should be resistant to microbial degradation and, therefore, is unlikely to be re-released to the atmosphere[48]. However, the perception that this material is not susceptible to degradation is increasingly challenged by

several lines of evidence. First, age poorly correlates with bioavailability[51–53]. Second, OM with terrestrial markers is found only in trace amounts in the ocean, suggesting degradability[54]. Common explanations for this paradox have included susceptibility to photodegradation, whereby photochemical alteration by short-wave radiation can enhance the cleavage of macromolecules into an array of low-molecular-weight (and bioavailable) compounds [55–57]. Similarly, microbial priming[54], where bioavailable C sources can trigger the degradation of previously unavailable organic matter[58,59], has been proposed as a mechanism behind this phenomenon. Alternative explanations include preferential sedimentation [60] and the dilution effect, which states that DOM persists in aquatic environments because the concentrations of individual molecules become so low that they are no longer efficiently utilised by microbes[61]. Together, these ideas suggest that the terrestrial DOM itself may be intrinsically biodegradable, but that some condition of the environment prevents its degradation.

There continues to be debate regarding why some material persists in aquatic environments[62–64], but advancement could likely be made if the concept of terrestrial DOM is refined. Specifically, we need to consider DOM from terrestrial ecosystems as a complex source. Since we don't fully understand the processes that underlie this complexity, it is difficult to answer important questions about the role, fate, and lifespan of DOM once it enters aquatic ecosystems. Therefore, a better understanding of the diversity of molecular formulae composing terrestrial DOM will provide vital information that can be used to understand how land-use practices such as forestry might influence the balance of C that moves through inland waters. Further, the resolution of terrestrial source quality will inform research exploring how C is portioned into various aquatic end-points and whether this offsets C capture efforts on land.

1.4 The transfer of terrestrial DOM to waters

In the previous section, I described reasons for understanding terrestrial DOM as a complex source; in this section, I will describe how DOM makes its way from and through various terrestrial source materials into the aquatic network. DOM from land has been described as a cold-water extract of terrestrial material— a watershed tea of biomolecules[65]. However, the components that make up this tea are not only a function of the possible source material but also the form and length of the path the water takes.

Headwater streams link the terrestrial and aquatic C cycles, and the hydrological path from the landscape to headwaters is critical for forming the incoming DOM mixture.

Reconciling runoff chemistry with hillslope flow path dynamics has been a focus of research over the last decade[66–68]. Generally, flow along hillslopes and into headwaters is dominated by groundwater and supplemented by inputs from soil water and surficial flow during flushing events (after snowmelt or precipitation)[69–73]. Upper soil and litter horizons are most like terrestrial plant source material and contain aromatic-rich compounds such as lignins, tannins, polyphenols, and melanins[74,75]. With soil depth, the aromatic material is lost, and soil environments become increasingly mineral-rich. It follows that the DOM character in soil solutions is constrained by the material it makes contact with. Furthermore, because the majority of DOM in both surface soil and groundwater share a common source (infiltrating precipitation or snowmelt), the biogeochemical evolution of DOM characteristics should represent the evolution of flow paths over the range of residence times and water sources. Therefore, conceptualising DOM composition as a continuum from recently leached surface soil and litter DOM, to longer residence time groundwater DOM, is consistent with catchment hydrology and water residence time distributions[76]. This model is somewhat complicated by the involvement of microbial communities continuously transforming DOM[77]. It is also unclear how the lateral movement of DOM along hillslopes affects its chemistry. Thus, understanding DOM source areas in soils and the transport of DOM from hillslopes to streams will help us understand the composition of incoming DOM, from which models based on its likely fate can be developed.

1.5 Fate of terrestrial DOM once in waters

I have explained the various sources of terrestrial DOM and the pathways DOM can take through the landscape. This section will focus on the various fates of this terrestrial DOM once it reaches aquatic ecosystems.

Terrestrial DOM reaching aquatic ecosystems undergoes biotic and abiotic molecular re-working and is either incorporated into biomass[78], out-gassed[79], stored in lake sediment[80], or transported through rivers to the deep oceans[81]. Aquatic ecosystems rely heavily upon terrestrial DOM to fuel aquatic primary production and

support higher trophic levels[82]. Terrestrial OM derived from vegetation and soils accounts for between 20 and 85% of secondary production in various freshwater ecosystems[83,84]. Tanentzap et al. (2014) found that at least 34% of fish biomass in freshwater deltas is supported by terrestrial primary production, with subsidies increasing with forest cover[82]. This close connection between terrestrial and aquatic ecosystems suggests that disturbances to forest ecosystems influencing DOM pools can also alter aquatic ecosystem structure and function[85,86]. Thus, we need a mechanistic understanding of how DOM quality relates to fate as both a means to reduce adverse effects (such as loss of C to the atmosphere) and as a step toward positive ecosystem manipulations, including management actions.

The fate of DOM in aquatic ecosystems is thought to be influenced by its source material, microbial community composition, and the environmental conditions experienced upon transport[87]. One useful theoretical concept proposed for understanding the transformation of DOM from land to ocean is the River Continuum Concept. This concept states that the diversity of DOM compounds decreases from headwater streams to river mouths, with only the refractory compounds remaining after the rapid degradation of labile compounds[88]. Given new data, revisions to this concept have been proposed. Creed et al.[89] suggested that DOM diversity increases with stream order as higher spatial and temporal diversity is incorporated, i.e., the River as a Chemostat Model[88]. Recently, Zark & Dittmar[28] introduced the idea of two levels of molecular diversity: β and α (similar to the ecological concept). In this case, β-diversity refers to the molecular DOM diversity between ecosystems. The α-diversity, on the other hand, refers to the molecular DOM species richness within an ecosystem. When this concept is applied to the stream-ocean continuum, headwater streams are expected to have high β-diversity reflecting the location-specific source material and microbial community. In contrast, oceans will have higher α-diversity as they integrate molecules (and isomers of the same structural molecules) from a large source area along the degradation continuum. It is unclear if this concept has a terrestrial counterpart or if similar processes apply to DOM in soil flowpaths (hillslopes) towards streams.

1.6 Advances in characterising the molecular composition of DOM

In the previous sections, I have outlined the sources of terrestrial DOM, the pathways of DOM from land to water, and the various fates of terrestrial DOM once it reaches aquatic ecosystems. These fates are believed to be driven by characteristics of the DOM mixture, which until recently, were not well characterised. This section overviews available methodology and the analytical advantages of ultra-high resolution mass spectrometry in resolving individual compounds in a DOM mixture and what molecular-level resolution of DOM can reveal about ecosystem ecology.

DOM is most routinely described by the concentration of bulk DOM[90]. In addition, techniques that consider the optical properties of DOM, such as UV-VIS sprectrophotometry[91] and fluorescence sprectophotomerty[92] are also commonly employed to characterise and quantify compound classes within the DOM mixture such as the amount of humic-like material or protein-like material. Advanced statistical approaches such as excitation-emission matrix (EEM) fluorescence spectroscopy with Parallel Factor Analysis (PARAFAC)[93] also allow for further characterisation into finer compound classes. Quantitative methods that also characterise the constituent of DOM also include size-exclusion chromatography coupled with organic C, organic nitrogen or UV detectors[94] and some advanced physical methods such as soft X-ray spectroscopy[95]. These techniques provide useful information but lack the analytical power to resolve specific molecules. There are also targeted analysis techniques of specific molecules available, for example of pesticides, that employ gas or liquid chromatography paired with mass spectrometry[96]. However, these routes do not enable high-resolution information about the entire DOM mixture.

Characterising DOM in terms of the individual molecules that make up its molecular composition is a powerful tool to gain fundamental insight into the fate, transformation, and source of DOM in aquatic ecosystems. However, the individual molecules and their variety of origins and ecological functions remain elusive due to: a) the extreme heterogeneity of DOM (up to 20,000 compounds can be found in a single sample[97,98]); b) the difficulty of complete dissolution (many organic compounds are insoluble)[99,100]; c) the lack of proper molecular separation from the ion matrix in which it exists in natural environments[99]; and d) the tendency of association in

complex superstructures when the characteristics of DOM emerge based on associations among compounds[101,102].

Despite these limitations, approaches resulting in a high level of molecular specificity often yield an impressive and robust set of conclusions that help develop novel insight into DOM processing[103]. In particular, there has been a remarkable progress in and increased use of analytical techniques such as Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS)[104]. Although still limited (as it cannot identify molecular structure[105,106] and de-salting is required[107]), this technique makes it possible to determine molecular formulas within the DOM pool with increasingly high accuracy and resolution[108] in a defined analytical window[98,109].

The wealth of information provided by FT-ICR-MS provides an opportunity to probe deeply into the complex interactions between DOM source, transformations, and fate. Although living organisms are obviously different from non-living organic compounds, a small but growing number of studies have applied concepts from ecological theory to the study of FT-ICR-MS- derived DOM data which shares a similar structure to biological species-abundance data[110–113]. Given the complex and abundant ecological data collected by early theoretical ecologists, these ecological theories relied on applications of systems theory[114] and information theory[115]. These applications followed two lines: first, to quantify the distribution of stocks and numbers of organisms, and second, to quantify the patterns of interactions between different levels of ecological networks (such as food webs). A similar framework has emerged in recent years for the study of ultra-high-resolution molecular data in ecology. Namely, the “ecology of molecules”[62] where individual compounds of the non-living organic matter pool and individual organisms interact within an environment.

Given the infancy of this discipline, the majority of the work developed in the ecology of molecules has been in the quantification and description of molecular stocks, with a few emerging examples of network anlayses[116,117]. For example, it is now known that natural DOM compounds consist of C (49.5±3.3%), H (5.0±1.0%), O (43.0±4.1%), N (1.7±1.0%), S (2.0±1.3%) and other elements[104]. Additionally, 13C-nuclear magnetic resonance spectroscopy data (capable of resolving DOM structure[118]) reveals that environmental DOM contains approximately 30% aromatic

groups, 23% alkyl groups, 22% carboxyl groups, 20% alkoxy groups, and 5% carbonyl groups[119]. FT-ICR-MS data can also provide a wealth of additional information related to the chemical features of DOM, for example, double bond equivalents (the degree of unsaturation) and aromaticity indices (unequivocal criteria of either aromatic or condensed aromatic structures)[120], the Kendrick mass (reveals groups of compounds that differ only by their number of CH_2 groups)[121]. DOM molecules can also be classified into lipid-, peptide-, amino sugar-, carbohydrate-, and nucleotide-like compounds according to their stoichiometric ratios[122], which can be visualised with graphical methods that sort compounds into regions based on ratios of H to C and O to C, such as van Krevelen plots[123]. Metrics of molecular lability have also been proposed, which partition data into more and less labile materials[124] based on their H:C ratios, or C, H and O presence[125]. These descriptors are an ideal starting point for understanding the role of particular molecular formulae in the environment, for example, why some molecules are rapidly broken down while others persist for centuries[62,63,126].

In addition to using chemical classifications to infer interactions with microbes and the environment, there has been increased interest in the ecological role of the diversity of molecules (a.k.a chemodiversity)[110,111]. Specifically, this line of inquiry treats biodiversity indices (a concept adapted from information theory[127]) as species where each formula corresponds to an individual of the species. Then, a relative species abundance, i.e., each species' fraction of the total individuals in a sample, is a measure of each molecule's abundance (e.g., signal intensity from FT-ICR-MS[127]). When paired with microbial diversity data, interesting questions emerge regarding the links between microbial and molecular diversity and composition (and microbes' ability to promote molecular diversity and vice-versa)[110]. Finally, spatial patterns in the processes in the ecology of molecules, have been increasingly applied across ecosystems, especially across defined spatial gradients[128]. These techniques borrow from the metapopulation paradigm and metacommunity theory[129,130] in addition to the field of biogeography. These tools from the ecology of molecules approach are well-suited for investigation of the land-water spatial patterns in DOM that may be altered by harvest and influence C budgets and aquatic ecosystem health.

1.7 Applying a molecular perspective to understand harvest impacts

A molecular perspective provides a new frontier from which forest harvest influences on aquatic ecosystem C cycling can be probed. With current priorities and technologies, economists now predict that the demand for wood products will increase by at least 19% in the coming decade[131]. As we explore and implement nature-based climate solutions[11], it is vital that we understand the wide range of impacts that forest harvest and climate change have on aquatic ecosystems via their link to terrestrial ecosystem dynamics. These impacts can be understood by investigating the molecular features of DOM[132,133].

Logging disturbances in forested catchments are generally known to produce long-lasting (many year) increases in the concentration of DOM exported into streams by altering hydrology[19,134,135]. Clearcutting and site preparation has been demonstrated to result in as much as a 19-35% increase in DOM concentrations and a 195% increase in riverine C flux in the first year post-treatment in northern Sweden[19]. In forests with seasonal snow cover, studies of harvest impacts on stream solutes and DOM concentration[135] found that when impacts occurred, stream DOM concentration typically recovered within three to six years. However, the impacts can persist for at least a decade after harvest[134,136–138]. One of the primary mechanistic drivers of increased DOM export after vegetation loss is increased runoff generation, i.e., when the water table rises, there is a more lateral flow of saturated soil water to the stream from superficial soil layers[67,139]. Soil depths up to 1m are estimated to contain 74% of all terrestrial C[16,140,141]. Thus, forestry may influence DOC mobilization by re-routing water flows that connect streams with different DOM source areas (upland soils, upper soil horizons, riparian zones). This observation was also supported by similar observations of greater proportions of streamflow generated from shallow flowpaths from a forest harvest study initiated in 1997 in the eastern boreal transition zone of Ontario, Canada[142,143]. Given the differential disruption of molecules throughout soil profiles, this will likely impact the molecular composition of DOM entering streams. This hypothesis has been tested a small number of times with optical approaches[144], but not with molecule-level resolution.

The effect of harvest is likely very complex, with changes to watersheds extending beyond hydrological ones. For example, logging disturbance is also a function of harvesting approaches, the volume of trees removed, the method and timing of tree removal, the area of the cut, and whether riparian buffers are retained[145]. There is also evidence that some forestry management practices displace more C than others (and likely from disparate places), for example, road building or driving heavy machinery into waterlogged soils[146,147]. Disturbing wet soils within and near stream and lake buffers can lead to increased surface flow through increased rutting and compaction. Moore *et al.* 2005 found that heavy forestry machinery traffic during forest harvesting changes water flow paths and increases soil wetness and temperatures[148]. Furthermore, forest harvesting decreases canopy interception and evapotranspiration, resulting in more snowpack water storage, soil moisture, and stream flow[149]. It can also shift canopy and understory species compositions and their development for more than a century[150,151] and alter potential wildfire behaviour and susceptibility to insect outbreaks[152]. Furthermore, harvest can alter soil microbial communities[153] and root exudate production[154], which can stimulate the turnover of C stored in mineral soils[155]. Thus, understanding the mechanisms controlling different magnitudes and duration of forest harvest responses will include understanding the nutrient demands of post-harvest re-growth and the altered physical and chemical environment within soils (which control solute production and mobility), in addition to the altered hydrologic flow paths[156]. This is a complex mixture of dynamic DOM sources whose contributions to streams are unlikely to be resolved by optical approaches alone. Unraveling the complexity of soil sources to streams is much more likely to be resolved utilising ultra-high resolution molecular information, which offers the promise of uncovering unique molecular tracers for land-use change[157].

The magnitude and significance of the influence of harvest practices on DOM quantity and quality have also yet to be reported in synoptic, landscape-scale studies. Previous research has demonstrated that the molecular composition of terrestrial DOM reflects complex ecosystem activities and is, therefore, highly diverse among ecosystems[47,158–161]. Meanwhile, there has been some evidence to suggest that within Swedish ecozones of the same type, there is a high similarity in the DOM molecular composition exported in headwater streams[162]. These results do

not *necessarily* contradict the earlier ones, but instead suggest that the relative proportions of molecules present could be predicted by ecozone rather than land-cover differences. However, an additional fluorescence spectroscopy study of boreal streams spanning multiple seasons found that land-cover type (wetland or forest) was the dominant factor (49% of variability) controlling DOM composition. Hydrological controls were found to be minor (8% of variability), and in-stream processing of DOM was unimportant[163]. Given the remaining uncertainty in the drivers of DOM composition on a landscape scale, as well as a paucity of harvest-specific studies, this is fertile ground for further investigation.

1.8 Thesis aims

This thesis addresses how forest harvest alters the quality of DOM exported to aquatic ecosystems from forested lands. By designing and implementing field experiments and synoptic surveys in the Batchewana watershed, Ontario, Canada, we were able to answer the following questions:

i. Does the molecular composition of DOM change similarly through soil depth and along hillslopes with different environmental conditions and, thus, suggest a potentially universal degradation process? (**Chapter 2**)

ii. How does harvest disturbance to soils affect the ecology of molecules in streams? (**Chapter 3**)

iii. How important is harvest relative to other landscape factors influencing DOM molecular composition, and do the harvest impacts on molecules in streams scale to the landscape? (**Chapter 4**)

In **Chapter 2**, we designed and established four replicate experimental headwater catchments in the Canadian hardwood forest. We tracked DOM along soil depths and positions on hillslopes in forest headwater catchments, relating its composition to soil microbiomes and physical chemistry to establish baseline conditions before a harvest experiment. Specifically, we considered how DOM changed along soil-aquatic gradients when considering the effects of environmental conditions and microbial re-working. Shared or general patterns of degradation can be used to predict the composition of DOM and how it will react downstream along

environmental gradients. This information can help manage C loss from soil to streams and understand the efficacy of nature-based climate solutions like forest harvest.

In **Chapter 3**, we used our experimental catchments from **Chapter 2** and performed a harvest treatment on two of the catchments, retaining two as controls. We then measured stream DOM molecular composition over the course of three years to measure and describe the effects of logging on DOM composition. We aimed to answer how the molecular composition of DOM changes, how the amounts of DOM exported from streams compared to the amount removed during forest harvest as wood, and how long these changes persisted. We further investigated the likely soil source areas using ultrahigh-resolution mass spectrometry measurements of both streams and soils. By working at a catchment scale, we can build a base of knowledge about how harvest disrupts aquatic ecosystems and their potential to release terrestrial C into the atmosphere.

In **Chapter 4**, we connected our understanding of terrestrial disturbance to aquatic ecosystems by examining the chemical properties of DOM in stream water from over 180 Canadian headwater streams in an area with historical forest harvest. We further utilised over 4500 fluorescence spectroscopy samples from headwater catchments (streams, soil pore water, and throughfall) to generate a representative fluorescence PARAFAC model for forested headwaters. We then examined the ability of the model to decompose key DOM properties by identifying associations with molecular classes from FT-ICR-MS measurements collected in **Chapters 2 and 3**. We subsequently examined how well components of the PARAFAC model and the inferred relationship to molecular information resolved by FT-ICR-MS could be explained by variation in the landscape and forest harvest. Thus, we first tested the efficacy of fluorescence spectroscopy and PARAFAC modelling, when validated with UHR-MS techniques, as a tool for monitoring the response of freshwater ecosystems to forest change. Second, we probed the relative importance of harvest impacts compared to other landscape controls of DOM molecular composition. Together, our results form a novel contribution to existing work on landscape-scale DOM composition by showing how simple, rapid, and cost-efficient PARAFAC models provide valuable proxy information on the molecular composition in streams when aimed at detecting the response of forested streams to harvest impacts.

Finally, I discuss the main findings of the thesis and possible avenues for future research in **Chapter 5**.

Chapter 2

Universal microbial reworking of dissolved organic matter along soil gradients

2.1 Abstract

Soils are losing increasing amounts of carbon annually to freshwaters as dissolved organic matter (DOM), which, if degraded, can offset their carbon sink capacity. DOM is more susceptible to degradation closer to its source and becomes increasingly dominated by the same (i.e., universal), difficult-to-degrade compounds as degradation proceeds. However, the processes underlying DOM degradation across environments are poorly understood. Here we found DOM changed similarly along soil-aquatic gradients irrespective of differences in environmental conditions. Using ultra-high-resolution mass spectrometry, we tracked DOM along soil depths and hillslope positions in forest headwater catchments and related its composition to soil microbiomes and physico-chemical conditions. Along depths and hillslopes, carbohydrate-like and unsaturated hydrocarbon-like compounds increased in abundance-weighted mass, suggestive of microbial reworking of plant material. More than half of the variation in abundance of these compounds was related to the expression of genes essential for degrading plant-derived carbohydrates. Our results implicate continuous microbial reworking in shifting DOM towards universal compounds in soils. By synthesising data from the land-to-ocean continuum, we suggest these processes generalise across ecosystems and spatiotemporal scales. Such general degradation patterns can be leveraged to predict DOM composition and its downstream reactivity along environmental gradients to inform management of soil-to-stream carbon losses.

2.2 Introduction

The fate of carbon exported from soils into aquatic ecosystems is a poorly understood component of the global carbon cycle. Soils store at least twice as much carbon as the atmosphere[164] and are expected to absorb more than one-third of anthropogenic emissions[165]. However, >15% of the net carbon added to soils annually from decomposing plant litter and roots is leached into aquatic systems as dissolved organic matter (DOM)[166]. Once in water, much of the DOM pool is highly reactive[30], potentially returning large quantities of carbon to the atmosphere as CO_2 or CH_4 and offsetting terrestrial carbon sequestration[23].

The distance from its terrestrial sources can predict how much DOM degrades in aquatic ecosystems[89]. Terrestrial DOM is initially dominated by a few spatially heterogeneous biomolecules, such as lignin-derived polyphenols that reflect local plant species composition[167]. Continuous transformation and remineralisation of this DOM along merging flowpaths produce an increasingly homogeneous pool of compounds downstream[88,168,169]. Compounds with structural features such as carboxylic-rich alicyclic moieties, material derived from linear terpenoids, and carotenoid degradation products dominate this increasingly homogenous pool[170]. As these compounds occur everywhere[28], that is, in all samples, they are termed "universal" or "core"[171]. The convergence towards a DOM pool dominated by universal compounds is consequently known as a "degradation cascade"[28]. Later stages along a degradation cascade should also have a greater proportion of shared compounds[28], such as measured with molecular β-diversity, should increase along this cascade[28], but this idea remains untested. All else being equal, the amount of time DOM is exposed to microbial and photochemical processing[62] is likely a unifying explanation for the degradation cascade[126]. Residence time correlates with carbon decay rates in marine sediments[172,173], bioassays[174], and inland waters[30,168]. As many universal compounds are degraded slowly by microbes[175]. DOM pools that are homogenised later along the degradation cascade can ultimately provide a persistent carbon store.

Universal DOM pools can result from common synthetic pathways or a chain of similar degradation steps[28,89,103,176,177], but how the importance of these mechanisms changes along diverse environmental gradients remains unknown[126]. Unlike along

depth profiles in the oceans, processing of DOM during vertical soil passage does not consistently converge to low-molecular-weight, recalcitrant compounds[178]. Degradation of DOM is instead characterised by increasing molecular weight. The process generating this increase in weight likely reflects the extracellular microbial decomposition of large macromolecular plant material into smaller molecules, for example, the production of simple unsaturated oligogalacturonates after pectin degradation[179]. These small metabolites are then transformed to larger microbially-derived compounds, namely complex polysaccharides associated with microbial tissues and products[177,178]., such as chitin or glucans[181]. Many microbes synthesise carbohydrates, hydrocarbons, and lipids in this way[182]. This process partly reflects the function of semipermeable cell membranes, where diffusion is restricted to molecules of low molecular weight, but these low molecular weight substances are subsequently elongated and incorporated into larger cell structures[182]. Degradation also varies with compound concentration[183], abiotic processes like sorption and desorption to minerals[184], hydrological pathways[185], and microbial trait diversity and energy supply[186]. Thus, DOM composition, microbial metabolism, and environmental and ecosystem properties interact to stabilise carbon[187]. Understanding how these processes influence DOM degradation along different flowpaths is necessary to ensure land-based carbon sequestration efforts are not offset downstream.

Here we asked if the molecular composition of DOM changes similarly through soil depth and along hillslopes with different environmental conditions and thus potential degradation processes. We worked across four replicate headwater catchments in northwestern Ontario, Canada (Fig. A1). We advanced previous studies by focusing on soil-water flow upstream of the headwater-ocean continuum, i.e., a natural but neglected extension of the riverine continuum[187]. We tracked DOM from 5 to 60 cm soil depth at each of shoulder, back, toe, and foot hillslope positions and into streams using Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) (Fig. 2.1a). We paired FT-ICR-MS with shotgun metatranscriptomic sequencing and metabolic measurements to reconstruct the function of microbial communities. Our fully factorial depth-by-hillslope design allowed us to test how DOM originating from the same source material changed along contrasting environmental gradients and into headwater streams.

Although we expected DOM to become increasingly homogenous across both soil depth and hillslope, consistent with a degradation cascade[28], we tested two contrasting predictions for how homogenisation occurred and the underlying mechanisms. The first prediction was that the extent of homogenisation and underlying physico-chemical conditions of microbial processes would differ. With depth, the DOM pool should reflect preferential consumption and transformation by microbial processes[178]. We expected DOM sources (i.e., plant litter) to be relatively consistent through depth with less hydrological mixing of sources than between hillslope positions. Along the hillslope, DOM should reflect different sources and hydrological mixing more than microbial processing due to gravity-driven differences in moisture, erosion, vegetation type and rooting depth. Therefore, we expected microbial processing along hillslopes should be relatively less important. Than with increasing soil depth because the relatively large variation in DOM sources. In contrast, our second prediction was that DOM composition would be homogenised similarly between these two spatial gradients of depth and hillslope. This pattern would suggest that universal processes, such asthe duration of microbial processing, shape similar DOM composition despite environmental differences along the degradation cascades. Our results now implicate athat common metabolic process shift DOM towards homogenised compounds along a soil-headwater continuum and suggest this process generalises across environmental and spatiotemporal gradients.

2.3 Results

2.3.1 Degradation of DOM across soil depth and hillslope

Consistent with the predicted increase in universal compounds from headwater to ocean[28], we found upland soils had fewer universal compounds than aquatic samples. We detected 12, 487 peaks and assigned 9327 unique molecular formulae (75% of peaks detected), more than twice that observed in a previous soil study, using the identical FT-ICR-MS method and instrument[178], suggesting we representatively sampled the DOM pool. We attributed this result to high extraction efficiencies for DOM (mean 69% ± 6% s.d.; see Methods), technical improvements in the detection cell, and advancements in molecular formula assignment[189]. Despite the many detected formulae, only 13% occurred in all samples, that is, were

universal, compared with between 47 to 87% in a synthesis from headwaters to oceans, using the same method and instrument[190]. Of the universal compounds, 79% were classified[123] as lignin-like based on the similarity of their elemental compositions to known biomolecules (Fig. 2.1b). Tannin-like (representing phenol derivatives) and condensed hydrocarbon-like compounds were the next most abundant classes based on counts of molecular formulae, accounting for 9%, and 6% of all formulae, respectively. Our results were not due to missing high-molecular-weight substances (HMWS) commonly detected in soils[191,192] but outside the analytical window of FT-ICR-MS[193]. Using liquid chromatography to quantify DOM fractions ,HMWS were detected in only 13% of soil pore water samples and, when found, contributed, on average (95% confidence interval, CI) only 8% (3 to 12%) to the total dissolved organic carbon concentration (Table A1).

Throughout the soil depth profile, the DOM pool converged upon a universal pool but not along landscape hillslope positions (Fig 2.1). Consistent with our predictions, the percentage of universal compounds increased from 20.9% (19.0 to 22.9%) of all formulae at 5 cm to 23.9% (21.9 to 26.1%) at 60 cm (Fig. 2.1c), generalised linear model: $z=2.1$, $p=0.036$, $df=57$). There was no change in the proportion of universal compounds between the shoulder position and streams (Fig. 2.1d; Table A2), as expected if hydrological mixing was important along the hillslope gradient. However, the DOM pool was similarly homogenised along both soil depth and hillslope gradients when we measured the proportion of signal intensity (i.e., relative abundance) attributed to universal compounds. The relative abundance of universal compounds increased by (95% CI) 9.4% (5.9 to 12.9%, linear model: $t=5.5$, $p <0.001$, $df=57$) from an estimated mean of 54.3% (51.8 to 56.8%) at 5 cm to 64% (61.4 to 66.0%) at 60 cm (Fig. 2.1e). There was a similar 8.1% (6.4 to 9.7%, $t=3.0$, $p=0.029$, $df=57$) relative increase in universal compounds from 56.9% (54.4 to 59.3%) at the shoulder to 59.8% (54.9 to 64.5%) at the stream, respectively (Fig. 2.1f).The relative abundance of universal compounds increased by (95% CI) 9.4% (5.9 to 12.9) and 8.1% (6.4 to 9.7%) from 5 cm to 60 cm and shoulder to stream, respectively (Fig. 2.1e,f). Universal compounds identified as those occurring in all our samples had similar molecular properties to literature definitions of degradation end-products that were independent of our sample set (Fig. A2). We found similar results when we matched our molecular formulae to those considered universal[194,195]

across aquatic ecosystems, again, likely because they reflect end-products of degradation (Table A2).

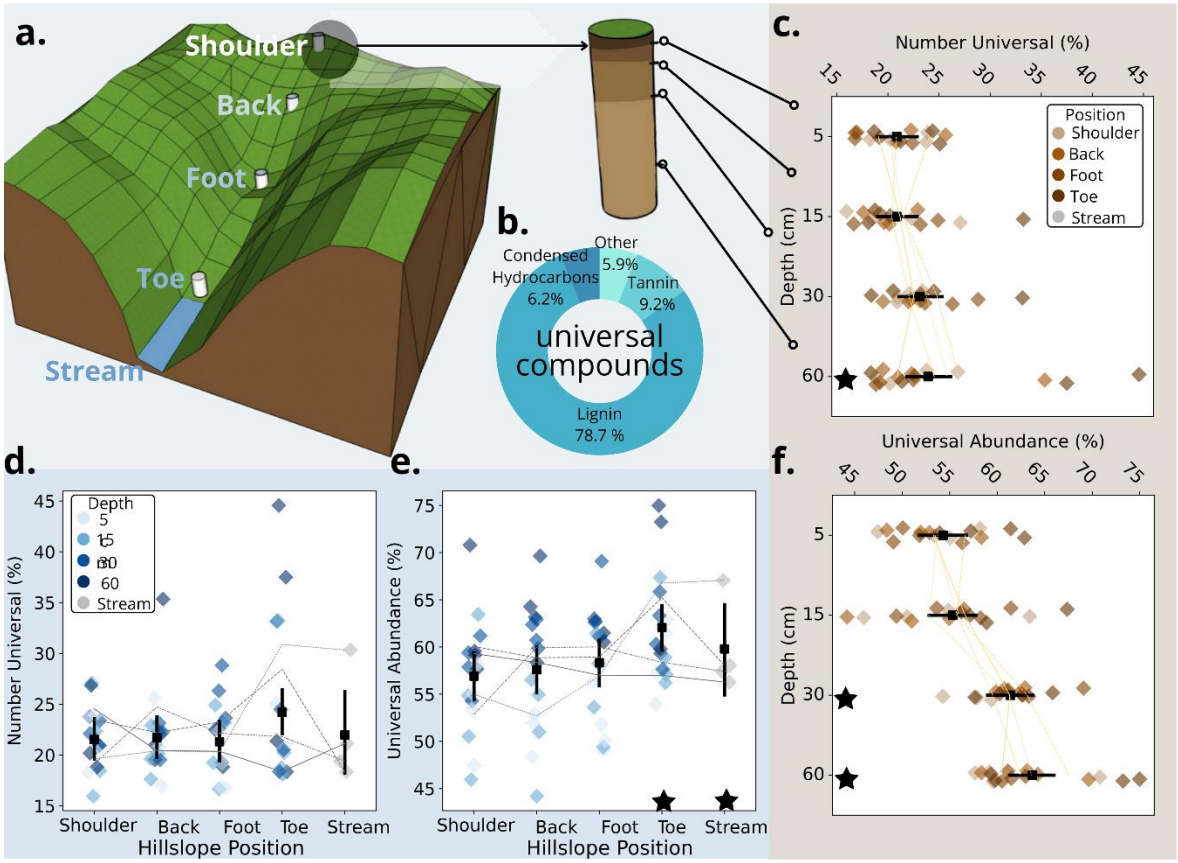


Fig. 2.1 DOM converges to a universal pool across soil depth and hillslope.Sampling design showing the location of soil lysimeters at shoulder, back, foot, toe hillslope positions and the headwater stream in a study catchment. At each hillslope position, we sampled soil at 5, 15, 30, and 60 cm depth. **b**. Relative contribution of classes to the universal compound pool (n=1216 molecular formulae). Lignins and tannins refer to molecular formulae that are phenol derivatives. **c**. to **f**. Composition of universal compounds. Points (diamonds) are percentage of compounds and relative abundance (i.e., sum of normalised signal intensities) comprised by universal compounds along the soil depth (**c.** and **f**., respectively) and hillslope (**d**. and **e**., respectively) gradients. Gradients of colour are increasing soil depth or hillslope position. Estimated marginal means ± 95% CI denoted by black squares were averaged across catchments and either depth or hillslope positions. Grey and brown dotted lines are mean values for catchment replicates for position and depth, respectively (n=4). Black stars denote hillslope positions and depths that are statistically different from either the shoulder

or 5 cm samples, respectively, based on Tukey-adjusted p values estimated from a generalised linear model (Table A2).

Convergence towards a universal DOM pool was due to the preferential degradation of subsets of compounds actively reworked by microbes. To understand better the biogeochemical processes driving the shift towards universal DOM, we deconstructed universal compounds into putative compound classes along both soil profile and hillslope gradients (Table A3). Universal lignin-like compounds (representing phenol derivatives) increased from an estimated mean (95% CI) of 40.1% (37.4 to 42.8%) of the relative abundance of all molecular formulae to 52.9% (50.2 to 55.6%, linear model: $t=7.8$, $p<0.001$, $df=57$) from 5 to 60 cm depth. These same compounds increased from an estimated mean of 43.6% (40.9 to 46.2%) to 48.4% (43.1 to 53.8%, $t=4.8$, $p<0.001$, $df=57$) from the shoulder position into the stream (Table A3, Fig. A4). The declines in relative abundance of universal molecular formulae from other compound classes were not large enough to account entirely for the increased representation of universal lignin-like compounds (Table A3, Fig. A4). Other non-universal compounds must have also become proportionally less abundant, such as if they were preferentially removed, for plant-derived lignin compounds to become increasingly represented in the DOM pool.

To explain further the shift towards universal compounds across the two gradients, we examined the relative abundance and intensity-weighted mass of non-universal compound classes. The relative abundance of non-universal tannin-like and condensed hydrocarbon-like compounds together declined by an estimated mean (95% CI) of 12.9% (8.9 to 16.9%, linear model: $t=5.5$, $p<0.001$, $df=57$) from 23.6% (16.9 to 20.8%) to 10.7% (7.9 to 13.5%) and by 13.3% (9.93 to 16.6%, $t=4.0$, $p=0.002$, $df=57$) from 21.2% (18.4 to 15.2%) to 14.7% (9.1 to 20.2%) along the depth and hillslope gradients, respectively (Table A4, Fig. A5). Both these compound classes tend to reflect plant material rather than microbial products[196]. Although other non-universal classes showed small average increases (<4%; Table A4, Fig. A5), these could not balance the declines in tannin-like and condensed hydrocarbon-like compounds, as expected if plant material was generally being degraded. In support of the degradation of specific compound classes causing a shift towards universal DOM, we found that intensity-weighted mass of non-universal compounds

increased along the depth and hillslope gradients (Fig. 2.2). These changes were similar between gradients despite differences in environmental conditions. From 5 to 60 cm depths, carbohydrate-, unsaturated hydrocarbon-, and lipid-like compounds increased in weighted mass by an estimated mean (95% CI) of 19.5% (10.8 to 21.6%, t=6.0, p<0.001, df=57) from 374 Da (357 to 391 Da) to 447 Da (430 to 464 Da,), 14.2% (7.8 to 17.7%, t=5.2, p<0.001, df=57) from 253 Da (242 to 263 Da) to 289 (279 to 299 Da), and 10.6% (4.7 to 15.0%, t=3.9, p<0.001, df=57) from 290 Da (278 to 301 Da) to 321 Da (310 to 333 Da), respectively. Carbohydrate- and unsaturated hydrocarbon-like compounds also increased consistently by 11.9% (6.0 to 23.3%, t=3.4, p=0.001, df=57) from 377 Da (360 to 394 Da) to 442 Da (408 to 476 Da) and 8.7% (3.9 to 20.2%, t=3.0, p=0.005, df=57) from 252 Da (242 to 262 Da) to 274 (264 to 284 Da), respectively, across hillslope positions from shoulder into the stream (Table S5). Although lipid-like compounds also increased from shoulder into the stream, they did not do so across the other hillslope positions like the carbohydrate- and unsaturated hydrocarbon-like classes (Table A5), potentially reflecting the larger sizes of lipids produced by aquatic primary producers[197]. The strength of the intensity-weighted mass shifts of non-universal compounds were large enough to increase intensity-weighted mass of all compounds along both soil profile and hillslope gradients (Table A6, Supplementary Results).Together, these results suggested that either larger mass compounds were being conserved or that new, heavier compounds were being created from lighter precursors. The latter scenario was expected if microbes were processing plant-derived compounds by degrading them into smaller compounds that were subsequently incorporated into larger microbial products[178].

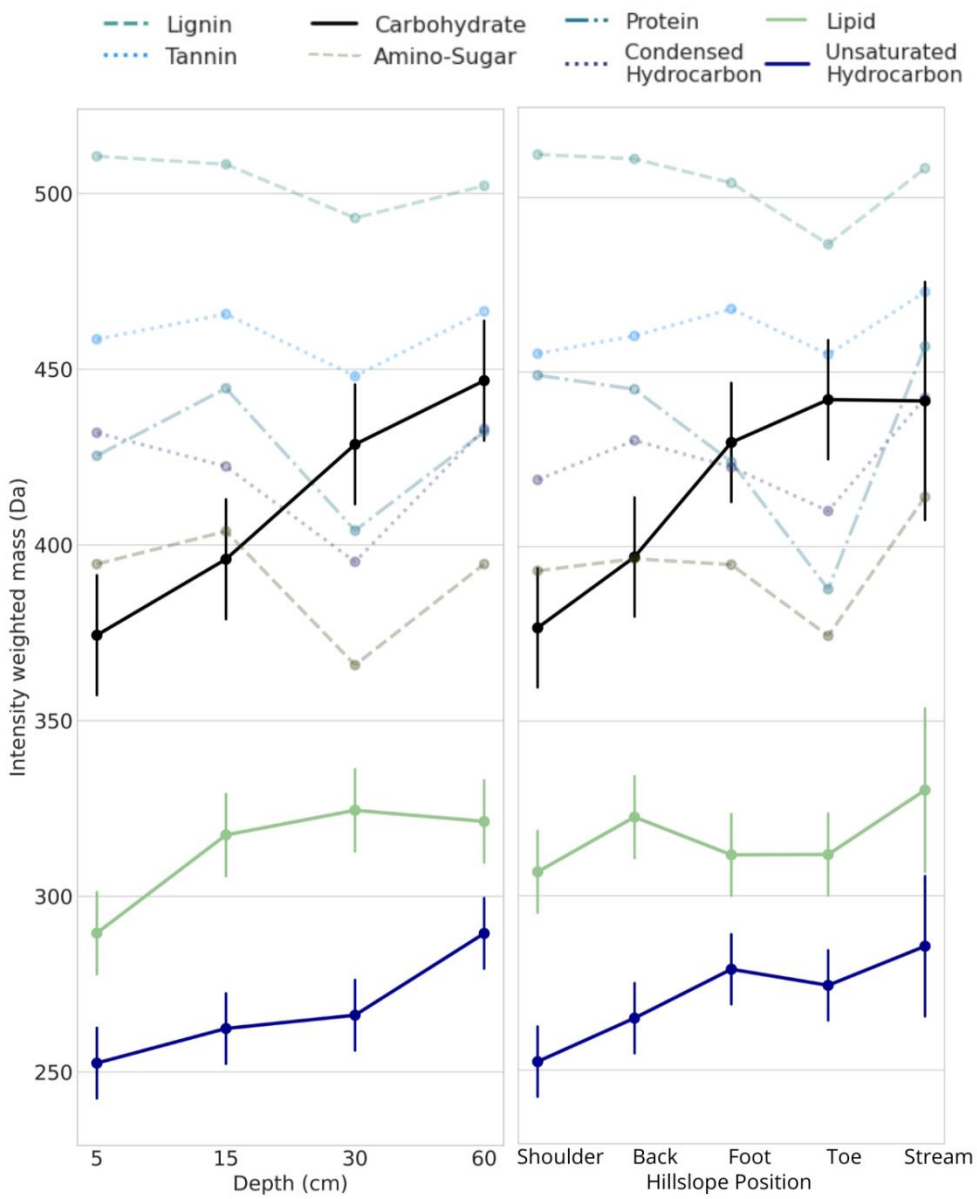


Fig. 2.2 Compound classes shift in mass across soil depth and hillslope position.Mean estimated molecular mass (± 95% CI) of DOM for each compound class when universal compounds were removed. Molecular formulae were grouped into compound classes based on their atomic ratios. Solid lines are those with statistically significant differences between 5 and 60 cm depths or shoulder and stream positions based on Tukey-adjusted p values estimated from a generalised linear model (Table A5). Errors for non-statistically-significant compound classes are presented in Table A5.

Environmental controls over the compound classes were identified by comparing soil depth and hillslope. As expected[199], absolute organic carbon concentrations declined from 5 to 60 cm and from shoulder to stream by an

estimated mean (95% CI) of 360% (274 to 474%, linear model: t=9.4, p<0.001, df=57) and 71% (66 to 87%, t=6.5, p<0.001, df=57), respectively (Fig. 2.3a). The decline in DOC was associated with greater microbial productivity along the hillslope (Fig. 2.3b; Table S7, t=7.2, p<0.001, df=25), suggesting that it could reflect microbial carbon consumption and not simply lower inputs of carbon at depth. Consistent with this result, we found a shift from humic-like to low-molecular weight, microbial-derived carbon along the hillslope using the larger analytical window of size-exclusion chromatography (Fig. 2.3c, Table A7). The ratio of humic-like substances to low-molecular-weight carbon chromatographic fractions decreased by an estimated mean of 57.7% (17.0 to 78.4%, t=2.6, p=0.013, df=57) towards the streams (Fig. 2.3c). By contrast, dissolved organic carbon to total nitrogen (C:N) concentrations in soil porewater decreased with depth, also as expected because of microbial processing[200], by an estimated mean of 119% (53 to 213%, t=4.4, p<0.001, n=57). There was no change in C:N ratios with hillslope position (Fig. 2.3d). Alongside the evidence of microbial consumption of dissolved organic carbon (Fig. 2.3a), nitrogen-rich proteins identified by FT-ICR-MS strongly declined from shoulder to toeslope (Fig. 2.2b). Nitrogen in the form of proteins was likely selectively adsorbed by clays[201] that accumulate at the bottom of hillslopes[202]. Thus, these results suggest that the environmental conditions differed between the depth and hillslope gradients (Fig. 2.3), yet the DOM pool consistently converged upon a universal compound pool (Fig. 2.2).

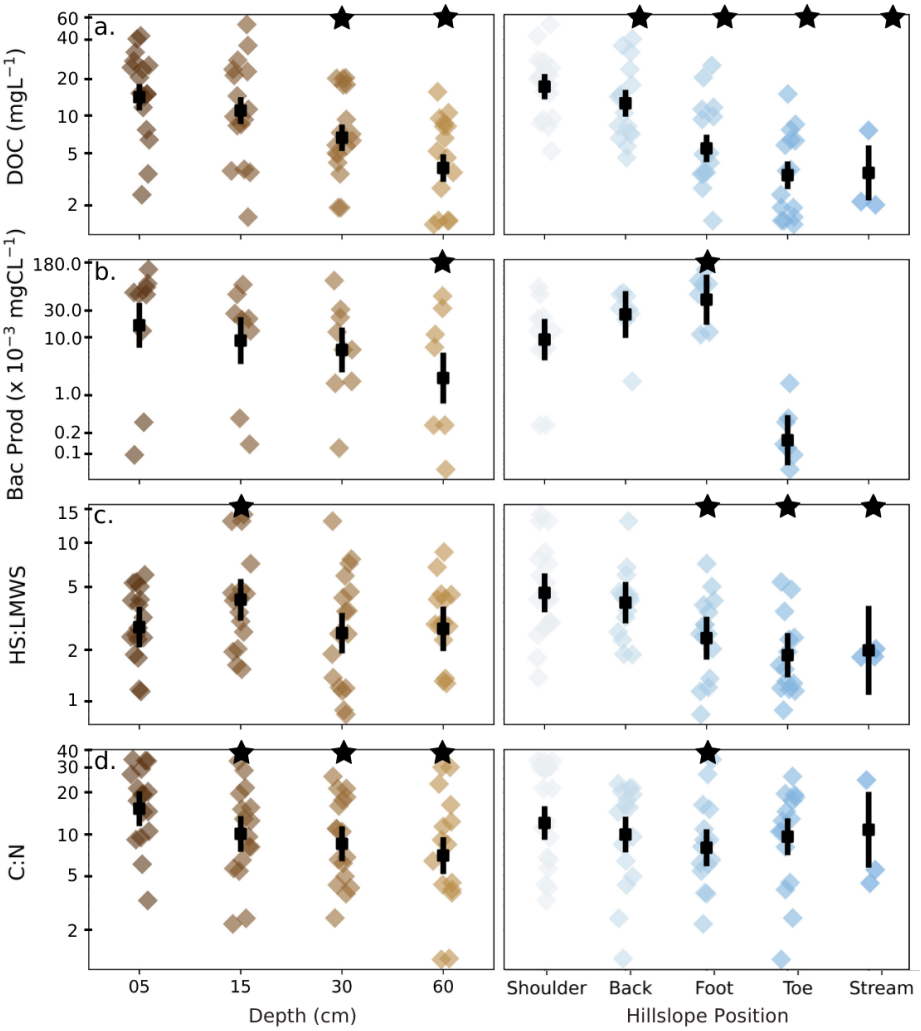


Fig. 2.3 Carbon concentration and quality change along soil depth and hillslope. We measured soil porewater for a. dissolved organic carbon (DOC) concentrations. b. bacterial protein production (Bac Prod; no stream sample was taken); c. fractions of humic substances (HS) to low molecular weight substances (LMWS) *;* and d. the ratio of total organic carbon to total nitrogen (C:N) concentrations. Points are estimated means ± 95% CIs. Black stars denote hillslope positions and depths that are statistically different from either the shoulder or 5 cm samples, respectively, based on Tukey-adjusted p values estimated from a generalised linear model (Table A7).

2.3.2 Microbial processing explain shifts towards a universal DOM pool

To understand why non-universal compound classes reflected increased microbial reworking, we partitioned the variation in their relative abundance by using redundancy analysis. We compared the importance of spatial variation along both

soil depth and hillslope position, including autocorrelation among sample locations, with environmental variables as sources of variation. We measured 475 environmental variables related to gene activity of carbohydrate-degrading enzymes, the activity of extracellular enzymes or level of carbohydrate substrate utilisation, and soil physico-chemical conditions (Table A8).To prevent overfitting, we reduced the environmental variables by removing highly inter-correlated parameters ($|r| > 0.80$) and then further selected predictor variables using an automated permutation-based procedure (see Methods). For carbohydrate-like and unsaturated hydrocarbon-like compounds that showed evidence of progressive reworking, that is, increased average weighted molecular mass (Fig. 2.2), we found that the spatial gradients (depth and position) were more important than for other compound classes (Fig. 2.4a). Overall, we likely identified many of the most important variables structuring DOM composition as our analysis explained 66 to 91% of the variation in the composition of compound classes (Fig 2.4b).

Most of the environmental variation in DOM composition was due to the potential and realised activity of microbial communities, particularly for compound classes that reflected increased processing along the degradation cascades. We explained differences in the processing of DOM by quantifying the contribution of each environmental variable to the variance partition analysis. This analysis identified 62 environmental variables that were important for explaining DOM composition across the 8 compound classes (Table A8). For carbohydrate-like and unsaturated hydrocarbon-like compounds, which showed increased reworking with both depth and hillslope, variables associated with realised microbial activity together explained more variation in composition than for any other compound class (25 and 27%, respectively; Fig. 2.4b).

Carbohydrate-active enzymes (CAZymes) primarily used to break down plant-derived carbohydrates were common when we identified the environmental variables shared only by compound classes that changed along the spatial gradients. For carbohydrate- and unsaturated hydrocarbon-like compounds, the activities of four CAZymes were exclusively shared in the lists of the most important environment variables (Fig. 2.4b). These CAZymes encoded lignocellulolytic enzymes involved in plant cell wall degradation (auxiliary activity 1[199]), glycoside hydrolases involved in degradation of both xylan (glycoside hydrolase 43) and microbial cell walls (glycoside

hydrolase 23)[199], and an enzyme involved in the breakdown of polysaccharide carbon-oxygen bonds (polysaccharide lyase 33)[201]. We also found evidence that these enzymes were expressed at a community-level as measured from enzyme activity assays. Catabolic use of two carbohydrate substrates (DL-α-glycerol phosphate and β-methyl-D-glucoside) by the microbial communities was also identified exclusively in the lists of the most important variables for carbohydrate- and unsaturated hydrocarbon-like compounds (Fig. 2.4b). For lipid-like compounds that also increased consistently in molecular mass with soil depth, only a family of cellulose-binding enzymes (carbohydrate-binding module 2[202]) was shared with carbohydrate- and unsaturated hydrocarbon-like compounds (Fig. 2.4b). These results provide more direct evidence than previously[178] that microbial processing underlies the progression towards a universal DOM pool along spatial gradients[48].

We further found that the soil microbiome shifted in activity from processing plant- to microbial-derived OM across the spatial gradients, providing among the first direct evidence of their importance in generating a universal DOM pool. We tested if the transcription of genes annotated as CAZymes were differentially expressed. We found that 17 out of 412 CAZymes statistically differed from 5 to 60 cm depth (Table A9). Of these, only polysaccharide lyase family 1, subfamily 2 (PL1_2) was also associated with classes that increased in molecular mass along the depth gradient (Fig. 2.4b). PL1_2 is involved in encoding pectate lyase that helps degrade plant-derived OM. As expected if OM inputs shifted from plant- to microbial-derived with increasing soil depth[178], expression of PL1_2 decreased from 5 to 60 cm by a mean (95% CI) of 5-times (2.0 to 11.0). By contrast, 61 CAZymes differed statistically from shoulder to toe positions (Table A10), of which 5 explained about 10%, on average, of the variation in compound classes that showed signs of processing with hillslope (Fig. 2.4b). The 5 CAZymes included auxiliary activity family 1, which was one of the most important for the composition of carbohydrate- and unsaturated hydrocarbon-like compounds (Fig. 2.4b), as expected if it was involved in oxidising phenols into these classes[186]. The other 4 CAZymes also all increased in expression by a mean of between 2.8 to 3.8-times (1.5 to 6.9) and were associated both with compound classes that increased (carbohydrate- and unsaturated hydrocarbon-like) and decreased (lignin- and protein-like) in molecular mass from shoulder to toe. These genes included PL1_2 and glycose hydrolases 13 (subfamily 18), 51, and 135, which are

associated with degradation of polysaccharides, including celluloses, and fungal biofilm and cell wall components. These results further implicate microbes in transforming plant-derived compounds fromshallower depths or higher hillslope positions into larger molecular weight compounds at later positions along degradation cascades.

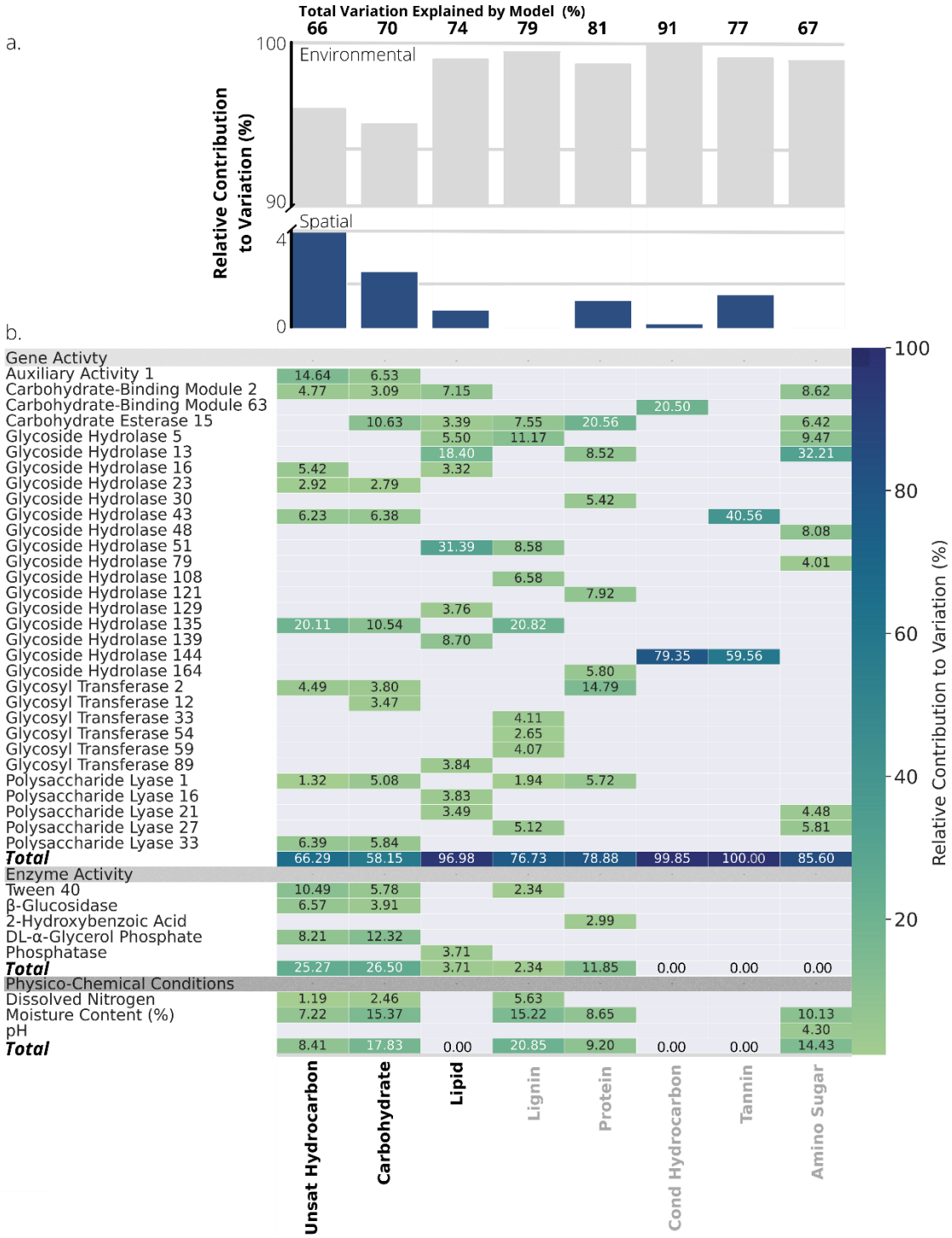


Fig. 2.4 Spatial gradients and environmental variables explain variation in the molecular composition of different compound classes. a. The relative variation explained by spatial and environmental variables in a variance partition analysis of the relative abundances of non-universal-molecular formula in each compound class. b. Above each bar is the absolute variation (%) explained by the model. Relative contribution of

environmental predictors in explaining total variance determined by hierarchical partitioning of analysis in a. Variables were grouped according to their association with the gene activity of carbohydrate-degrading enzymes, the activity of extracellular enzymes or level of carbohydrate substrate utilisation, and soil physico-chemical conditions. Variation explained by a gene family was displayed if >2.5% for at least one compound class and was summed across all identified subfamilies within a family. Bolded compound classes increased in molecular mass with depth and/or hillslope.

2.3.3 Generalising the persistence of DOM across the land-ocean continuum

Our study provides new evidence that DOM converges towards a universal compound pool as microbial reworking removes components typical of shallower soil depth and higher hillslope positions. To determine if this process could apply more generally, we contextualised our observed patterns within the headwater-ocean continuum by synthesising published FT-ICR-MS data. Rather than calculate universal compounds in a global pool of samples, we measured convergence as similarity to a deep-sea reference sample to facilitate inter-study comparison. Although the percentage of compounds shared among all samples in a study reflects convergence towards a universal pool, it is sensitive to differences among studies in molecule number, formula attributions, and sampling intensity. The deep-sea sample is the endpoint of degradation along the land-ocean continuum, so should accumulate the highest proportion of universal compounds[204]. We subsequently found samples expected to be exposed to microbial processing for longer, that is, further along the land-ocean continuum, such as deep soils and toeslope and stream positions, were most like deep-sea DOM (Fig. 2.5). The degree of convergence towards a universal pool was consistent with our observations for upland soils, generally, and the position of our different study depths and hillslopes (Fig. 2.5). These results suggest that a similar process of cumulative exposure to microbial processing may explain changes in DOM along similar spatial and temporal gradients, though the extent to which these trends are linear may vary with time[205], hydrological mixing[206], and rooting depth[178]. This process would provide the mechanistic basis for popular heuristic models like the river continuum concept[207] and soil chromatograph[203]. More sophisticated ecosystem models could now leverage this pattern in DOM composition to predict variation in DOM degradation[186] and identify potential carbon sinks[204].

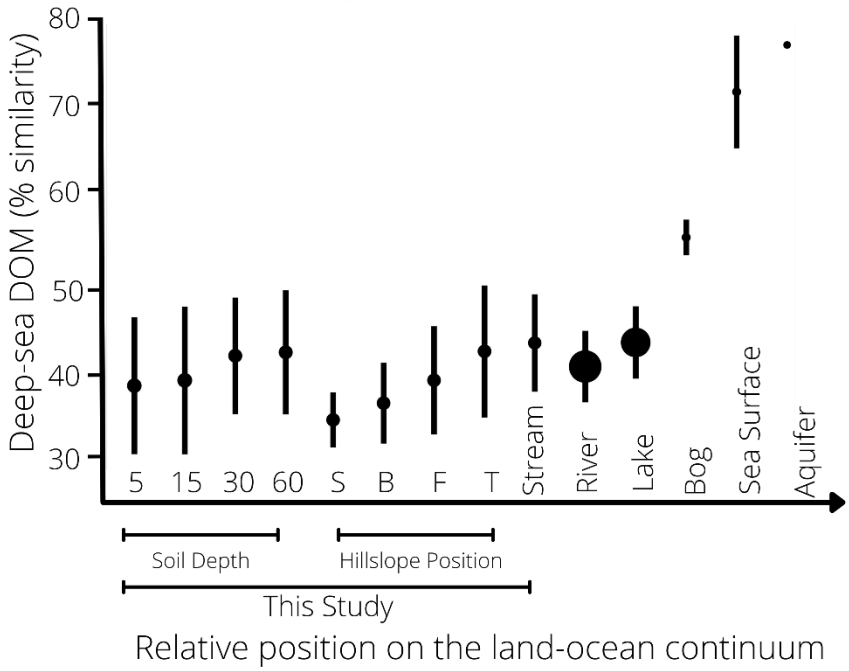


Fig. 2.5 ***Soil pore water is an extension of the aquatic continuum.***

The mean percentage of molecular formulae shared with a deep-sea reference sample was calculated for 433 samples from 5 published FT-ICR-MS studies across a headwater-ocean continuum (Table A11). To account for differences in per-sample molecule number arising from variation in instrument resolution and sample preparation, we used a rarefaction approach employed in microbial diversity studies[208]. We randomly sampled each study 1000 times to a set of 6000 compounds. This threshold was determined based on a minimum number of molecules to adequately sample the pool of observed compounds (Supplementary Methods). Points are means ± standard deviations of observations. Sizes of points scale with the number of observations at each position along the land-ocean continuum (n = 2 to 116). Results for soil depth and hillslope position were averaged across sites for each respective gradient.

2.4 Discussion

Because compounds vary in their persistence, by demonstrating that microbial processing changes the abundance of individual compounds along different environmental gradients, our work advances our understanding of how organic compounds accumulate and influence soil carbon sequestration. Two dominant processes likely contributed towards this result. First, universal compounds, especially those that were lignin-like, may have been selectively retained because of the energetic costs required for their degradation[209]. Second, non-universal compounds (for example, in the carbohydrate-like class) were consistently reworked by microbes. Genes encoding enzymes involved in the breakdown of plant cell walls were especially related to mass shifts of microbially produced compounds, e.g., carbohydrate-like rather than lignin- or tannin-like. Given the ubiquity of plant cell walls in nature, these results suggest widespread microbial metabolic pathways, like those involved in the production of pectate lyases[210], could underpin a general land-to-ocean degradation process. This process could persist independent of microbial taxonomic composition and the effects of environmental conditions on certain compound classes, e.g. mineral protection[211]. Even if environmental conditions were modulated along the land-ocean continuum by time or other, unmeasured environmental factors, the convergence toward a universal compound pool that we have observed across different environments implication a general degradation process. Our study was not designed to test a third hypothesis commonly invoked to explain DOM accumulation in oceans: compounds persist because their concentrations are too low to interact with microbes[61]. Despite homogenisation of chemical traits, DOM has been found to become more structurally diverse along the degradation cascade[28]. Individual structures may occur at concentrations that limit microbial activity[188], and our paired FT-ICR-MS, shotgun metatranscriptomic sequencing and metabolic methods approach could be applied in the future to test this idea explicitly. More generally, the reasons for the spatial variation in the degradation state of DOM that we identified here can improve soil-to-stream carbon management.

2.5 Methods

2.5.1 Field site and water sampling

Samples were collected from four forested catchments in northwestern Ontario, Canada (46° 58′ N, 84° 22′ W, altitude 375 m; Fig. A1). Within each catchment, hillslopes were partitioned into four positions: shoulder, backslope, footslope and toeslope. Classifications were based on the morphology of surface digital elevation models[208] and a Height Above the Nearest Drainage terrain model[209] that provides a spatial representation of soil-water environments[212]. At all hillslope positions, soil water was sampled in October 2019 after two months of continuous calibration sampling. Samples were collected at 15, 30 and 60 cm depths with tension lysimeters that consisted of 60-mm-long round bottom necked porous cups with an outer diameter of 48 mm and effective pore size of 1.3 µm (model 0653X01-B02M2, Soilmoisture Equipment Corp, USA). The sampling bottles were evacuated to a negative pressure of 50 kPa with a hand pump so suction pressure was ca. 50 mbar above the actual soil water tension. At 5 cm depths, lysimeters could not be securely installed. Therefore, we sampled pore water with micro-tensiometers designed to extract fluids non-destructively using a vacuum[189] through a 0.6 µm ceramic cup (Rhizon CSS samplers, Rhizosphere Research Products, The Netherlands). All the water samplers were installed in triplicate and pooled at each of the depth-by-position combinations to retrieve sufficient volume. Surface water from streams was grab-sampled at the bottom of each hillslope at the channel head. The hillside design of 16 soil samples and 1 stream sample was replicated across the four catchments for a total of 68 samples.

All water was filtered within 8 hours through 0.45 µm glass fibre filters (Whatman GF/F, pre-combusted 400°C, 4 h) and treated differently depending on the eventual analyses. For fluorescence spectroscopy, samples were aliquoted with no headspace into 20 mL borosilicate scintillation vials and stored at ca. 4°C in the dark. For LC-OCD, samples were frozen in acid-rinsed and pre-combusted 22 mL borosilicate vials with a PTFE/silicone septa and a polypropylene cap. For DOC concentrations and FT-ICR-MS, samples were acidified with 37% trace metal grade HCl to pH 2 and stored in pre-combusted 40 mL amber borosilicate vials with a PTFE/silicone septa and polypropylene cap at ca. 4°C in the dark.

2.5.2 DOM characterisation and concentration

To characterise DOM, an aliquot of acidified sample (0.1–30 mL, depending on the DOC concentration) was desalted and concentrated via solid-phase extraction (SPE) using a styrene-divinylbenzene copolymer sorbent (100 mg cartridges, Agilent Technologies, USA) using established methods[189]. The sorbent was pre-soaked in HPLC-grade methanol the night prior to extraction. Cartridges were sequentially rinsed with ultrapure water, methanol and ultrapure water acidified with HCl to pH 2. Acidified ultrapure water was stored in the same type of bottles as the procedural blanks. After loading the SPE cartridges with sample, the cartridges were rinsed with acidified ultrapure water and dried. The DOM extracts were eluted with methanol. On each day of extraction, a process blank extract was produced. Redundancy analysis confirmed that the extraction efficiency had no influence on the molecular composition of DOM (explained variation = 0.7%, pseudo-F = 0.6, P = 0.722).

We analysed the DOM extracts on a solariX FT-ICR-MS with a 15 Tesla magnet (Bruker Daltonics, Germany). The system was equipped with an electrospray ionisation source (ESI, Bruker Apollo II) applied in negative ionization mode. We diluted methanol extracts to a final concentration of 2.5 ppm in a 1:1 (v:v) methanol:water solution before injecting 100 µL into the FT-ICR-MS instrument. For each measurement, we collected 200 scans in duplicate. An in-house mass reference list was used for internal calibration using Data Analysis Software Version 4.0 SP4 (Bruker Daltonics, Germany). The mass error of the calibration was <0.06 ppm for all samples. All measurements were done within 7 days and in random order. In-house reference samples[213,214] were used to confirm instrumental stability.

Masses ranging from 150 to 1000 m/z were exported from the Bruker Data Analysis software, and we assigned molecular formulas using the online formula assignment and analysis tool ICBM-OCEAN[215]. A method detection limit of 2 was applied to all exported masses[216,217]. All molecules were singly charged ions and we limited formula attributions to $C_{0–100}$, $H_{2–200}$, O_{0-70}, $N_{0–3}$, $S_{0–2}$, $P_{0–2}$ with a tolerance of 0.2 ppm. Only signals detected in both duplicate measurements were retained. We normalised peak intensities of the peaks with an assigned molecular formula to the sum of peak intensities. Intensity-weighted mass-to-charge ratios (m/z_{wm}) for each

DOM sample were calculated as the sum of the product of the m/z_i for each compound *i* and their relative intensity I_i divided by the sum of all intensities $m/z_{wm} = \Sigma(m/z_i I_i)/\Sigma I_i$. We also assigned molecular categories to all formula using established criteria[218]. We further defined the structural features[219] in each category to ease semantic confusion among classification schemes (Fig. A3).

A LC-OCD system was also used to characterise the DOM pool, including compounds not captured by the analytical window of FT-ICR-MS. With size-exclusion chromatography coupled to a Gräntzel thin-film UV reactor, this procedure differentiated[220]: (i) non-humic "high molecular weight substances" (HMWS) of hydrophilic character; (ii) aromatic humic or "humic-like substances" (HS) with higher aromaticity based on UV absorbance at 254nm; and (iii) "low molecular-weight substances" (LMWS) including acidic and neutral substances. DOC fractions and molecular weights were determined based on compound retention times using the dominating HS peak in the carbon detector chromatograms to identify the position of other fractions. The position of the HS peak and molecular weight of the HS fraction were calibrated using International Humic Substance Society Suwannee River humic acid and fulvic acid standards[221].

2.5.3 Soil Sampling

Microbial activity was measured in soils, as pore sizes of the soil water sampling devices may have excluded soil-bound microbes contributing to the DOM pool. Soils were collected using manual shovelling at each hillslope position in two catchments. Three sterilised 30 cm PVC cores were hammered into an exposed soil wall at each sampling depth (5, 15, 30, and 60 cm). Soil from these cores was homogenised and sieved through a 6 mm mesh to exclude roots and inorganic material using sterile tools. A 10 g subset of soil was cooled on ice for bacterial productivity and respiration experiments performed within 24 hours of sampling. The remaining samples for bacterial and fungal taxonomic diversity (250mg) and CAZyme gene quantification and identification (2g) were stored in sterile centrifuge tubes or freezer bags and frozen on-site using dry ice in ethanol. Samples used for sequencing and microbial activity (enzyme assays and substrate use) were stored at -80°C and -20°C, respectively.

2.5.4 Environmental predictors of DOM

We measured 475 environmental variables in soils and their porewaters to partition variation in DOM composition. The variables were associated with microbial metabolic activity and biomass (n=38), microbial diversity (n=6), expression of CAZymes (n=412), and soil water chemical and physical properties (i.e., physico-chemical conditions: n=19) (see Table A8 for full list).

We measured 38 variables related to microbial activity and biomass. Bacterial production (BP) was measured using ^{3}H leucine incorporation after Bååth et al[222] (Supplementary Methods). Decays per minute measured were converted to BP (mg L^{-1} day^{-1}) using standard conversion factors[223]. For microbial basal respiration, 1 g of field-moist soil was placed in a 10 mL glass vial and incubated for 24 h in the dark at room temperature (21°C) without manipulating moisture levels. Incubations of live cells were initiated within 5 hours of soil collection. Respired CO_2 (µL min^{-1} g^{-1} dry soil) collected in the headspace of the vial was measured with an infrared gas analyser (QS102, Qubit Systems, Canada). We also measured the potential activity of four hydrolytic enzymes in cell suspensions: beta-xylosidase and beta-glucosidase that breakdown xylose and oligosaccharides, respectively, N-acetyl-B-D-glucosaminidase that degrades glycoside and amino sugars, and phosphatase that degrades proteins. All enzyme activities were assayed in 96-well plates under controlled conditions (pH 5, room temperature for 1-hour) using 4-methylumbelliferone-fluorescence tagged substrates and measured with a Synergy H1 Hybrid spectrophotometer/fluorometer (BioTek Instruments, USA) as in existing protocols[224]. To complement the enzyme assays, we measured microbial substrate use of 31 different carbon sources using Biolog® EcoPlates™ (Biolog Inc., USA). Cell suspensions were prepared by adding 10 g dry soil equivalent to 95 mL of sterile NaCl solution (0.85%), 6 ceramic beads and 5 g glass microspheres. Samples were mixed on an orbital shaker for 30 minutes at 200 rpm, left to settle for 15 minutes, and a serial dilution on 1 mL supernatant performed to an end concentration of 1:1000 in saline. Inoculated plates were incubated in the dark at 25°C, and absorbance read every 24 hours on the Synergy H1 microplate reader. Blank wells were always subtracted to reduce noise. Average well colour development (ACWD) was calculated as the rate of change over a 24 hr period from

day 1 to day 2, as day 2 produced the greatest ACWD indicating the least chance of substrate limitation. Finally, bacterial biomass was measured using flow cytometry. Bacterial cells were separated from soil matrices using buoyant density centrifugation adapted from ref. [225] (Supplementary Methods). Flow cytometry was performed on Accuri™ C6 Plus flow cytometer (BD Biosciences, UK) equipped with a 200 mW solid-state laser emitting light at 488 nm, measuring green fluorescence at 520 nm (FL1 channel). The FL1 and forward scatter detectors were used to reduce autofluorescence found in environmental samples. We ran samples in triplicate, passing 100 µL per technical replicate through the flow cytometer at a speed of 60 µL min^{-1} to prevent overlap of scatter events. Gates were set by comparing scatter plots produced from stained and unstained samples. Flow cytometer counts were validated with Spherotech 8-peak and 6-peak beads. Bacterial biomass (mg g^{-1} dry soil) was calculated from cell counts, assuming a conversion factor[226] of 58 fg $cell^{-1}$.

Bacterial and fungal taxonomic diversities were assessed using exact sequence variants (ESVs) generated by amplicon sequencing of the 16S rRNA gene and ITS2 region, respectively. DNA was extracted from 250 mg of homogenised soil using the DNeasy PowerSoil Pro kit and the QIAcube® (Qiagen, Germany) automated platform. 16S rRNA and ITS2 libraries were prepared following[227]with well-established primers (Table A12). The one exception was the first set of PCR reactions were set up by mixing 25 µL of HotStarTaq Plus Master Mix, 19 µL RNase-Free Water (Qiagen, homogenised), 0.5 µL of 10 µM primer and 5 µL of gDNA at 5 ng $µL^{-1}$. Indexed and purified amplicons were quantified using the Synergy™ Mx Microplate Reader (BioTek Instruments, USA) before pooling at equimolar concentration. Libraries were sequenced paired-end (2×250 bp) on the Illumina Miseq platform at the Aquatic and Crop Resource Development Research Centre, National Research Council Canada, Saskatoon at an average (±SE) read depth of 23594 (±863) and 34017 (±2380) reads for 16S and ITS, respectively. Sequence data were processed using the MetaWorks pipeline version 1.4.0. To reduce potential bias introduced by both large differences in read depth (i.e. >10-times difference) and small, uneven libraries, we removed samples with <1000 reads and remaining samples were rarefied to the 15th percentile of reads (7150 and 11103 for 16S and ITS, respectively) using the rrarefy function in vegan[221]. Eight 16S and six ITS samples with slightly less than the 15th percentile were also kept (≥6097 and

≥7544 reads for 16S and ITS, respectively) based on rarefaction curves that showed saturation. Reads were then taxonomically annotated with the RDP classifier v2.13 and UNITE classifier v2.0 for 16S and ITS, respectively. We calculated bacterial and fungal diversity with the Shannon-Weiner index that accounts for relative abundances of ESVs in addition to their number using the diversity function in the R package vegan[222].

To identify CAZyme genes and quantify their transcripts, we used metagenome and metatranscriptome shotgun sequencing, respectively. Shotgun metagenomic libraries were prepared with the Nextera XT DNA library preparation kit and the Nextera XT Index kit v2 (Illumina, USA) following the manufacturer's instructions using the same input DNA that was used for amplicon sequencing. DNA libraries were purified with Agencourt AMPure XP beads (Beckman Coulter, USA) and fragment size (250–1000 bp) verified on a 2100 Bioanalyzer with a high sensitivity DNA kit (Agilent, USA). Libraries were quantified with the Qubit BR dsDNA assay kit and pooled at equimolar concentrations prior to pair-end sequencing (2×150 bp) at the Centre d'Expertise et de Services Génome Québec on an Illumina Novaseq platform. Metatranscriptomes were obtained by extracting RNA from 2 g of soil using the RNeasy® PowerSoil Total RNA Kit (Qiagen, Germany), except that the phenol/chloroform step was repeated twice. The pellet was suspended in 50 µL RNase/DNase-free water, treated with the RNA Clean & Concentrator-5 with DNase I treatment kit (Zymo Research, USA), and eluted in 15 µL of DNase/RNase-free water. RNA quality was verified with the 2100 Bioanalyzer using the RNA 6000 Nano or Pico assay (Agilent, USA), while RNA concentration was determined with the Qubit RNA high-sensitivity assay kit (Life Technologies, USA). Absence of residual DNA in RNA extracts was further confirmed by PCR amplification of the 16S gene. rRNA was depleted from RNA extracts using the Pan-Prokaryote riboPOOL-kit (siTOOLs Biotech, Germany) with hydrophilic streptavidin magnetic beads (New England Biolabs, USA). rRNA-depleted RNA was then purified with the RNA Clean & Concentrator-5 kit and eluted into 10 µL of DNase/RNase-free water. Libraries were prepared using the NEBNext Ultra™ II RNA Library Prep Kit for Illumina (New England Biolabs, USA) and the NEBNext Multiplex Oligos for Illumina kit (New England Biolabs, USA) following the manufacturer's protocol for rRNA-depleted RNA. A quality check of the libraries was

performed on the 2100 Bioanalyzer with the high sensitivity DNA kit (Life Technologies, USA) prior to pooling and pair-end sequencing (2×125 bp) on an Illumina HiSeq platform at the Aquatic and Crop Resource Development Research Centre, National Research Council Canada, Saskatoon. Metagenomes were screened with Fastp[221] for read adaptor removal and co-assembled per sampling site with metaSpades[228] (V0.6.1) using Kbase[229] according to default parameters and including the BayesHammer option for read error correction[230]. Gene sequences were identified on the assembled contigs using Prodigal[231] and then annotated as CAZymes using Hidden Markov Models from dbCAN (V9[232], e-value < 1e-15; coverage > 0.35) in local searches with HMMER v3.1b1[233]. Metatranscriptomes were quality-filtered with Fastp according to default parameters[234] and mapped against gene sequences confirmed as CAZymes to obtain their expression profiles using CoverM (v0.6.1 using the 'tpm' option, https://github.com/wwood/CoverM). Transcript counts were then normalised using the R package DESeq2 to correct for library size and composition and allow for comparison between samples[131,235].

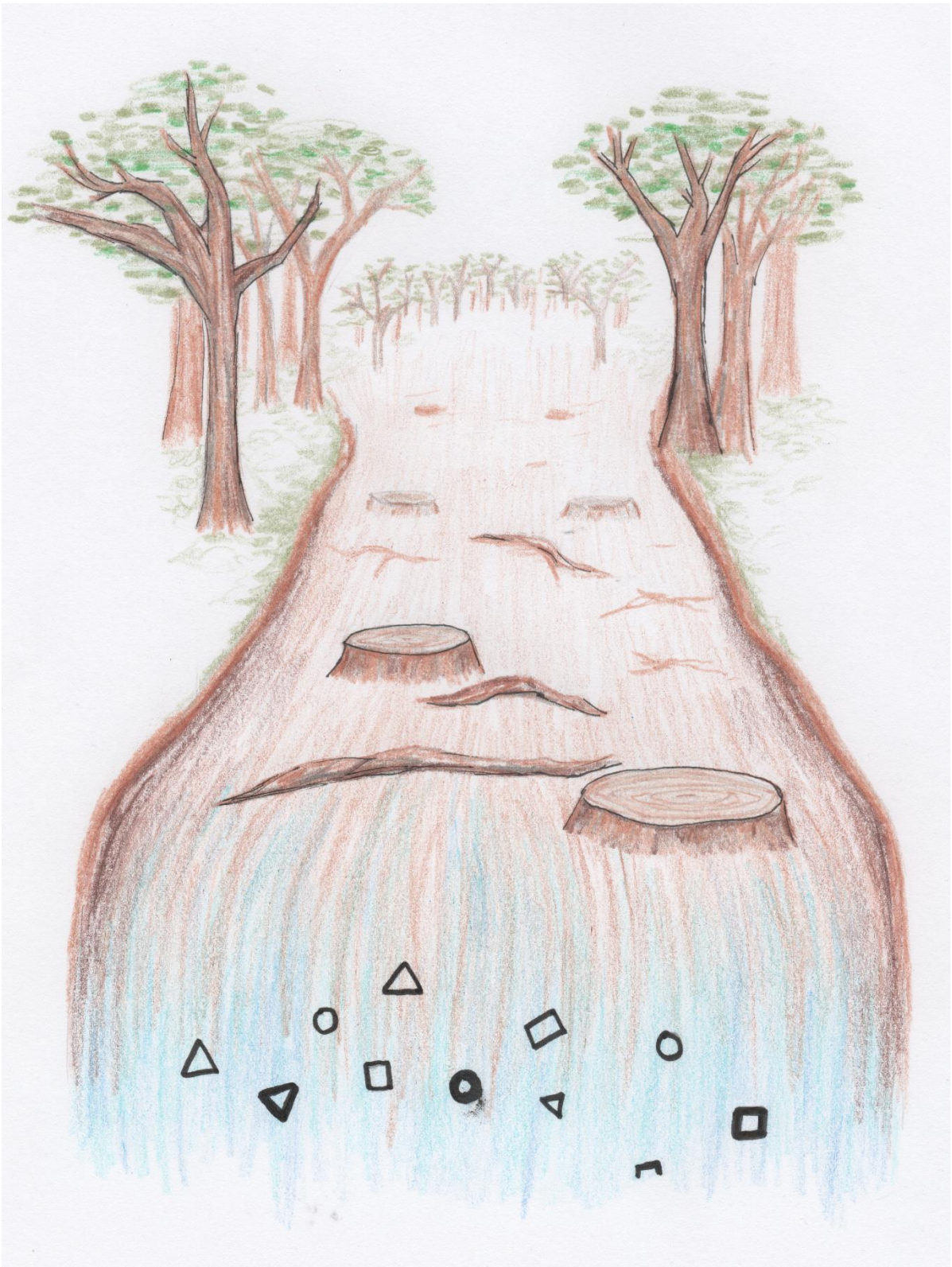
Finally, major ions, nutrients, and metal concentrations were measured from a subset of the soil pore water at the Great Lakes Forestry Centre, Sault Ste. Marie, Ontario, according to methods outlined in Table A13. Soil moisture content was directly measured by weighing the change in the mass of ca. 10 g of soil before and after drying in an oven for 24 hours at 105°C degrees relative to the original mass.

2.5.5 Statistical analysis

We tested if the molecular composition of DOM varied with soil depth and hillslope position using generalised linear models in R version 4.1.2. The probabilities of detecting universal molecules, percent relative abundance of universal molecules, and intensity-weighted mass-to-charge ratios (m/z_{wm}) were predicted at each soil depth and hillslope position, while also allowing responses to vary simply because of catchment identity and if the sample was soil or stream. We used a binomial error structure for models where proportions of counts were predicted and accounted for over-dispersion by including an observation-level random effect. Models were otherwise fitted with a Gaussian error structure and responses that were a proportion of continuous variables were log- or logit-transformed to normalise residuals.

Marginal means were predicted at each depth and position by averaging across catchments and either depths or positions using the R package emmeans[9,236].

To compare the importance of environmental and spatial drivers underlying DOM composition, we used a redundancy analysis (RDA) inferential framework[235,237]. We partitioned the total unique and shared variation in the relative abundance of each compound class explained by all environmental variables, spatial structure, and the depth and hillslope gradients using RDA ordination estimated with the varpart function from the vegan package[233]. Spatial structure was modelled using principal coordinates of neighbour matrices [238,239]; see Supplementary Methods. Prior to the variance partition analysis, the 475 environmental predictors were reduced to avoid overfitting. Where two or more variables were correlated with a Pearson $|r| > 0.80$, we removed the variable with the largest mean $|r|$ with all other variables. We further reduced this subset (n=364) for each compound class with stepwise model selection[235] using the ordistep function from the vegan package, dropping variables that were weakly associated ($p > 0.10$) with molecular composition. The importance of individual environmental predictors in each compound class were tested using hierarchical partitioning generalised to multiple predictor matrices implemented with the R package rdacca.hp[236]. All environmental predictors were scaled to zero mean and unit variance.



Art by: Riley Gladney-Hatcher

Chapter 3

Logging disrupts the ecology of molecules in headwater streams

3.1 Abstract

Globally, the demand for wood products is driving increased forest harvest. One potential and understudied consequence of deforestation is the mobilization and loss of organic matter from soils to rivers as dissolved organic matter (DOM). Here, we quantify and characterize the effects of logging on DOM with an experiment using four replicate catchments of the Canadian hemi-boreal hardwood forest over three years. DOM concentration in stream water from logged catchments increased as a pulse during the first year, but only changes in quality persisted thereafter. Ultrahigh-resolution mass spectrometry revealed that DOM released from logged catchments was more energy-rich and chemically diverse, with novel additions of highly-unsaturated polyphenols, carboxylic-rich alicyclic, and nitrogen-containing formulae. Given the molecular composition of the streams post-harvest, we suggest that the source of diverse DOM was due to higher hydrological connectivity with intermediate and deep soil layers, contrary to previous models. We estimate that while logging increased the overall annual flux of dissolved organic carbon by approximately 8.5% of the extracted wood carbon, the exposure of deeper soil horizons through logging released previously stable soil organic carbon to streams where it may be differentially available for microbial degradation. The resulting changes to the molecular composition of DOM within headwater streams persisted for at least 2 years after logging, potentially disrupting aquatic ecosystems and making streams more likely to release terrestrial carbon into the atmosphere.

3.2 Introduction

Economists predict that the demand for wood products will increase by 19% in the coming decade[20,240], especially with the shift toward natural climate change solutions[9,238]. Harvest of wood from plantations and secondary and primary forests must rise to meet this demand[237,239], but may have unintended environmental consequences. Forest removal and soil disturbance associated with wood harvest typically increase dissolved organic matter (DOM) export to aquatic ecosystems[19,134] though decreases have been reported[241,] especially in tropical systems[85,242,243]. Even without considering harvest-associated increases in DOM export, at least 15% of annual terrestrial net ecosystem productivity is exported from soils into inland waters, which is conservative, as this estimate only includes the contribution of mineral soils[177,244.] Once in waters, DOM is highly reactive[124,245,246] and more likely to be partitioned for respiration (rather than microbial growth)[177,244], making carbon much more susceptible to re-release to the atmosphere from waters than in soils. Therefore, the quantity and composition of DOM exported from soils into waters may offset the intended benefits of natural climate change solutions.

The effect of forest harvest on DOM export and composition remains a major unknown in global C budgeting. DOM is a complex mixture of compounds, where each compound, due to its intrinsic properties and its interactions with the environment and microbes, has a unique likelihood of being assimilated into microbial biomass or respired to the atmosphere as CO_2 or CH_4[247,248]. These outcomes can now be somewhat resolved for individual molecular formulae in this complex mixture using ultra-high-resolution mass spectroscopy (UHR-MS). Communities comprised of molecular formulae can specifically be explored under the emerging framework of the “ecology of molecules”. This framework studies how individual compounds (or molecular formulae) of the non-living organic matter pool interact with all other compounds and microbes in a particular environmental context[249–251] . Key features of DOM that predict its reactivity under the ecology of molecules framework are the number of unique molecular formulae in a mixture (hereafter called “chemodiversity”) and the proportion of different compound classes, typically identified based on elemental ratios of individual formulae[243]. Previous studies suggest that both chemodiversity and compound composition are tied to the

fate of aquatic carbon[252–254]. For example, carboxylic-rich alicyclic molecules (CRAMs) containing relatively intermediate H:C and O:C ratios tend to be less reactive over time and so may persist better in the environment [255], whereas high H:C (>1.5) aliphatic, protein-like, lipid-like, and high N and P-containing formulae[256,257] are preferentially degraded by microbes and so may be more easily respired[258]. If increased wood production is to aid the shift towards low carbon societies, there is a need to understand better how logging impacts the ecology of molecules, especially via chemical features tied to reactivity.

Historically, aquatic chemodiversity sourced from terrestrial ecosystems was assumed to be resistant to biodegradation and therefore escape atmospheric release[249,250]. However, this view has been revised given evidence that old and terrestrial DOM is rapidly oxidized by microbes once released into aquatic ecosystems[259,260]. With the understanding that DOM mobilised after harvest may be biologically active, disturbance that mobilises deeper soil carbon into downstream waters risks introducing otherwise stable C into the pool of reactive C and increasing the likelihood of release to the atmosphere. Ultimately, whether compounds are degraded, flocculate into larger mixtures, or are deposited into sediment depends on DOM composition[261]. Preliminary investigations show that human disturbance to forests, in addition to changing DOM fluxes, exports DOM from deeper soils[262], however, little is known about the mechanism driving the altered patterns in DOM composition in streams after forest harvest. Changed DOM composition could indicate both disturbed soils and altered flowpaths[260]. Knowledge of this mechanism is essential to design effective forest management practices that protect aquatic ecosystems and minimize the potential of downstream carbon emissions.

First-order streams are ideal to investigate the mechanisms driving DOM composition after land-use change because they carry a stronger imprint of their surrounding watersheds than higher order waterways[263–265] and their terrestrial inputs are simpler to resolve. Forest harvest is thought to increase DOM export[184] in these streams by first modulating the factors controlling soil solution concentrations of DOM[249–251] (such as temperature[263], moisture[263,264], logging residues[262]) and second by increasing lateral groundwater flow from deeper to more superficial soils rich in DOM[266]. DOM composition and chemodiversity in streams may be impacted by the same processes that govern DOM export. Namely, if hydrology changes, so

too should the soil layers that water interacts with before entering streams. Soil DOM can become homogenised along soil depth and hillslope position gradients[267], and so should respond similarly to water that interacts with deeper soil layers after harvest. Furthermore, in river networks, Lynch at al.[184] observed a loss in DOM chemodiversity with increasing flow conditions because of reduced physical opportunities for microbial metabolism. Given the analogous increase in flow under post-harvest conditions, logging could produce a similar homogenizing effect on DOM composition, i,e. lower chemodiversity.

Here, we experimentally tested how DOM composition and chemodiversity changes after forest harvest in first-order streams and surrounding hillslope soils between 2019 and 2021. We replicated our study in four hardwood-dominated catchments, two of which were harvested in 2020 with paired controls selected for nearly identical climate regimes, underlying geology, size, and topography (see Methods). We quantified spatial and temporal differences in DOM molecular composition before and after forest harvest using a before-after-control-impact statistical design and complementary analytical chemistry approaches, namely Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) and fluorescence spectroscopy. We report large changes in DOM composition after forest harvest and their likely sources via changes to soil flow paths, thereby identifying levers for forest management aimed at climate change mitigation[268.]

3.3 Results

3.3.1 The immediate mobilisation of DOM after forest harvest

Consistent with the expectation[269,270] that forest harvest increases DOM export to streams, we found DOM concentrations were elevated in the logged catchments relative to the controls in the year of harvest (Fig 3.1b, Table B1). This selective forest harvest (see Methods) largely offset the decline in DOM seen in the controls after increased seasonal discharge[271–273] (Fig 3.1b). The rate of change in concentrations in the harvested streams (0.08 mg C L^{-1} d^{-1}; SE: ± 0.31 mg C L^{-1} d^{-1}) was large enough that the mean DOM concentration would double from initial values (1.22 ± 0.38 mg C L^{-1} d^{-1}) every 30 days relative to the controls (Fig. 3.1b). Cation

and heavy metal concentrations increased similarly in the streams of the harvest sites, indicative of increased nutrient leaching from disturbed soils[52,274,275] (Table B2). To understand how the DOM fluxes compared to estimates of carbon removed as wood biomass by logging, we estimated annual carbon export for the harvested catchments using the modelled DOM increase and monthly runoff in 9 similarly-sized local catchments (see Methods). We estimated an additional mean 13 kg C $month^{-1}$ ha^{-1} (SE: ± 4 kg C $month^{-1}$ ha^{-1}) of DOM exported from the logged sites compared to the non-harvested catchments. Assuming this monthly mean DOM increase remained constant for a year, an additional ca. 8.5% of the C harvested as above-ground woody biomass during logging would be lost (see Methods).

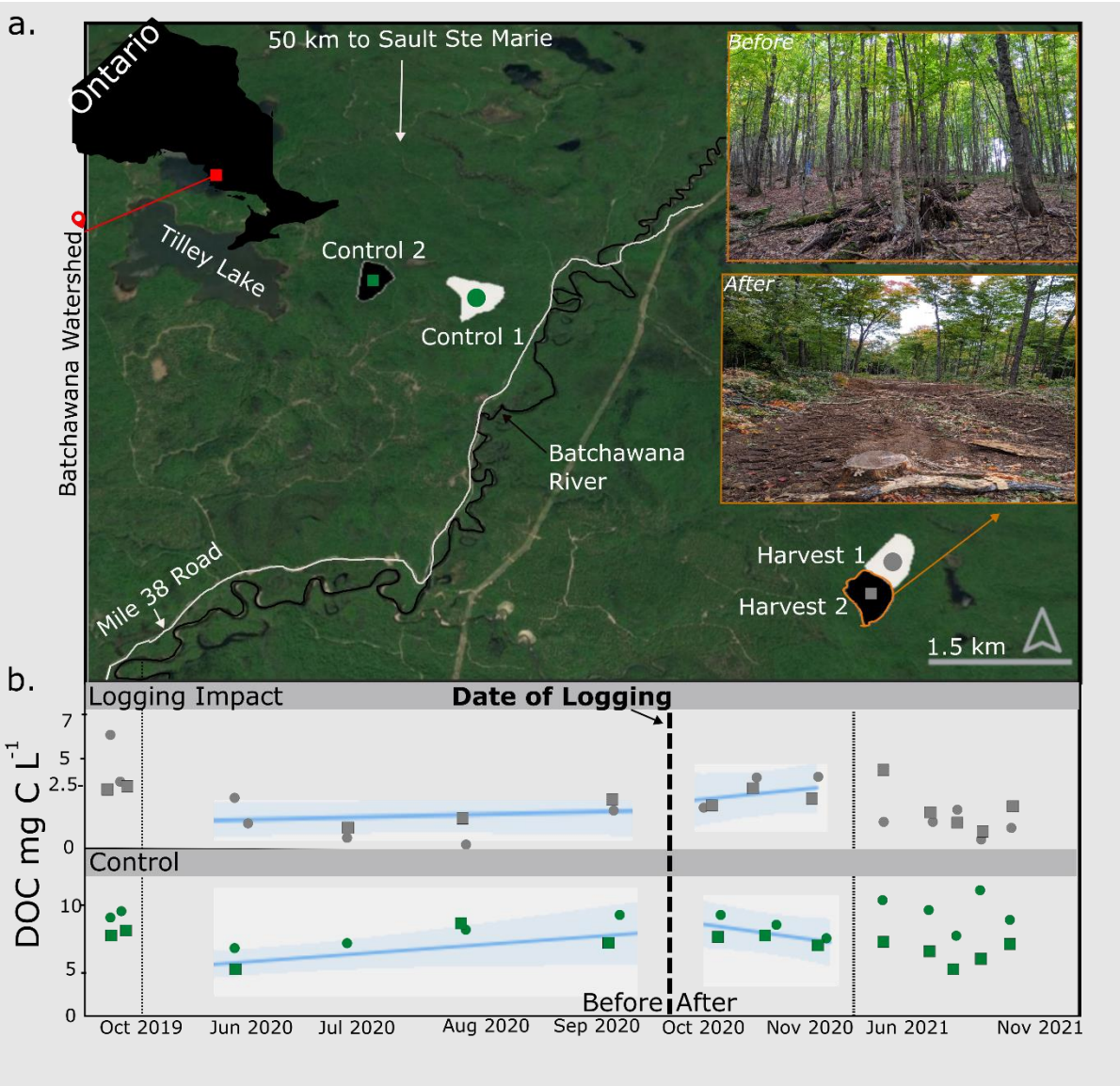


Fig. 3.1 DOM concentrations were elevated in streams of the harvest relative to the control sites immediately after logging.a. Map of experimental catchments within the

Batchewana Watershed, Ontario, Canada. The four replicate catchments were named based on the impact classification. Control sites were unharvested, and Harvest sites were logged in late September 2020 (bolded dashed line in bottom panel). **b.** Faceted plot showing dissolved organic matter (DOM) concentrations categorised by period (Before, After), year (2019, 2020, 2021) and treatment (Harvest upper, Control lower). Coloured circles and squares correspond to individual catchments as indicated in panel a. Estimated mean ± 95% confidence interval for temporal trends were plotted as blue lines ± bands where there were statistically significant differences between treatments (Table B1). Seasonal variability in discharge and precipitation can be found in Fig. B.1.

3.3.2 Changes in DOM composition persist after forest harvest

Our results suggest that while there is an increase in DOM concentration only in the year of harvest, the molecular composition of DOM in streams remained influenced by forest harvest well after the initial DOM pulse (Fig. 3.1b; Table B3). The humification index (HIX), which is a unitless measure of the complexity of DOM, increased by 22% from a mean of 4.58 (SE: 1.01) before the harvest to 5.58 (SE: 0.92) within 2 months after harvest. This result indicated that DOM was becoming more polycondensed (lower H:C ratio), or, more soil-like[274]. The increase in HIX was a mean 3.80 (SE: 1.75) larger than in the controls between 2019 to 2021. In contrast to the increase post-harvest, HIX declined between 2019-2021 by a mean of 17% in the control (SE: 1.2) from a mean 16.05 (SE: 1.27) to a mean 13.24 (SE: 0.92), indicating long-term leaching of soil-DOM only into the harvested streams.

Using UHR-MS of solid-phase extractable DOM, we deconstructed the fluorescence signal into individual molecular formulae. We detected 7444 distinct molecular formulae across all streams sampled in 2021. While optical indices reflect only the mean or bulk character of DOM[274], we identified a specific fraction of the overall molecular formulae (13 and 16%) that were positively and negatively correlated with HIX, respectively (Fig. 3.2a). We then correlated commonly used descriptive traits of individual molecular formulae to the strength of the correlation between the relative intensity of each formulae and the HIX value of the corresponding DOM pool across all sampling times. Molecular formulae that were more positively associated with HIX also had larger values of the formula-based estimate of aromaticity (AI_{mod}, $\rho = 0.62$, $p < 0.001$), which indicates the presence of

condensed aromatic structures[184] and confirms that stream DOM was becoming more soil-like post-logging. Studies in lake sediment have found that AI_{mod} strongly correlates with the relative abundance of lignin-like compounds[184], further supporting a terrestrial plant origin to the HIX signal rather than changes produced in-stream[158,276]. The association with AI_{mod} was stronger than with any other metric (Fig 3.2b).

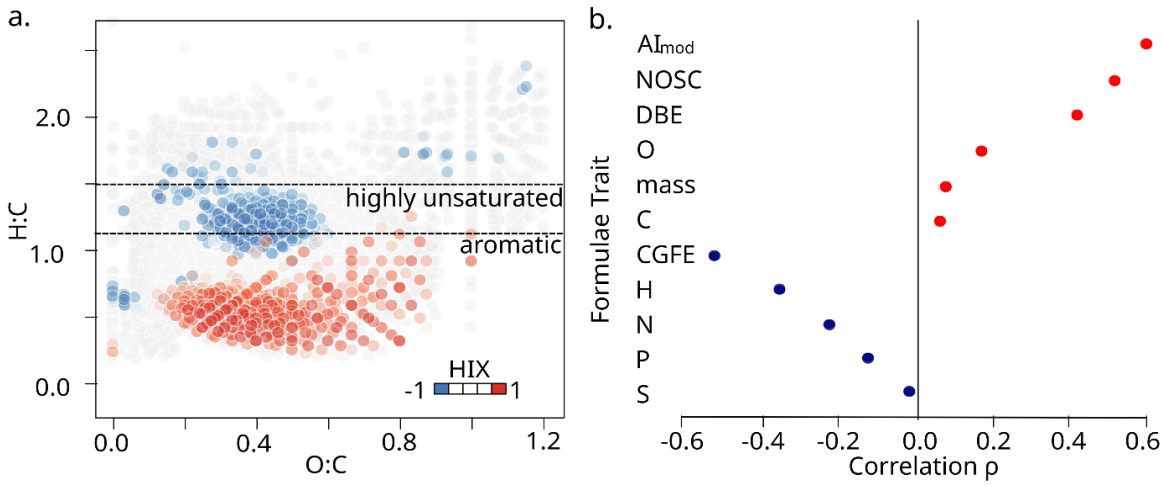


Fig. 3.2 Forest harvest led to persistent increases in the humification index (HIX) of DOM in headwater streams.a. Elemental ratios of all molecular formulae detected in our experiment (n=7444). Red and blue points are formulae whose relative peak intensities correlate ($p<0.05$) positively and negatively with HIX, respectively, based on Spearman rank correlations. Dotted lines indicate aromatic compounds (H:C< 1.1) and the "highly unsaturated" region (1.1 < H:C < 1.5) according to ref[54]. **b**. Spearman rank correlation (ρ) between either the modified aromaticity index (AI_{mod}), nominal oxidation state of carbon (NOSC), double bond equivalent (DBE), number of oxygen atoms (O), molecular mass, number of carbon (C) atoms, Gibb's free energy of carbon oxidation (CGFE), and the number of hydrogen (H), nitrogen (N), phosphorus (P) and sulphur (S) atoms of each molecular formula against the correlation between the relative abundance of each molecular formulae and the HIX of the corresponding DOM pool in across all time points (n = 7) measured with UHR-MS. For all correlations, $p < 0.05$.

3.3.3 Increased diversity of harvest-impacted stream DOM

We expected that forest harvest would reduce chemodiversity in streams as DOM became homogenised by high soil hydrological connectivity induced by harvest[255].

However, we found a 32% increase in the number of unique compounds after forest harvest compared to the controls (Fig 3.3a). The number of unique compounds increased from a mean of 4771 (SE: 330 compounds) to 6292 (SE: 390 compounds), whereas compounds declined in controls from a mean of 5387 (SE: 399 compounds) to 4153 (SE: 260 compounds). We found further support for this diversification post-logging using a multivariate analysis of molecular composition. We calculated the Bray-Curtis dissimilarity between control and harvest sites at each time point. The pairwise distance between paired sites increased after harvest (Fig. 3.3b), suggesting that compound communities were increasingly differentiated. Furthermore, the harvest sites differed more between the before and after samples than the controls (Fig. 3.3). These results are the first direct report of the diversification of DOM molecules after forest harvest, and contrast expectations based on observational studies[184] that the compound community should become homogenised.

We explored the mechanism underlying the unexpected increase in chemodiversity by analysing the molecular and elemental composition of compounds that were gained in the logged sites after forest harvest. Generally, vascular plant-derived DOM is aromatic and has a low H:C and high O:C ratio[278–281]. Consistent with this interpretation, we found that the compounds gained after logging had higher O:C ratios (Fig. 3.4c). There was also a greater proportion of highly unsaturated phenolic oxygen-rich compounds, i.e., compounds enriched in aromatic structures (Fig. 3.4a), as expected if there was a greater contribution of surficial soil layers and harvest residues to the DOM pool. We also found evidence for contributions from deeper soil layers. On average, N-containing DOM from the logged catchments in this study comprised 43% of the newly added compounds, but only 32% of all identified compounds (Fig. 3.4b, Tables B7, B8). Most of the added N-compounds fell outside of the H:C range interpreted as being associated with autochthonous inputs from primary production[282] (Fig. 3.4a). Instead, higher N content likely corresponded with release of aged organic carbon[20] through disturbance of deeper soils. The concomitant increase in carboxyl-rich alicyclic molecules (CRAMs) supports this hypothesis, as the relative intensity of CRAMs increases with microbial processing of carbon and is known to be greater in deeper soils[283] (Table B8). Together, the input of highly unsaturated phenolic oxygen-rich compounds, N-

containing formulae, and CRAMs suggest logging increases chemodiversity downstream by disrupting both surficial and deep stores of soil carbon. This result builds on the surface contribution model[196], often posed as an explanation for increased DOM concentrations[284] after harvest. Our results demonstrate that alongside these shallower flowpaths, deeper soils also make contributions that are activated by disturbance.

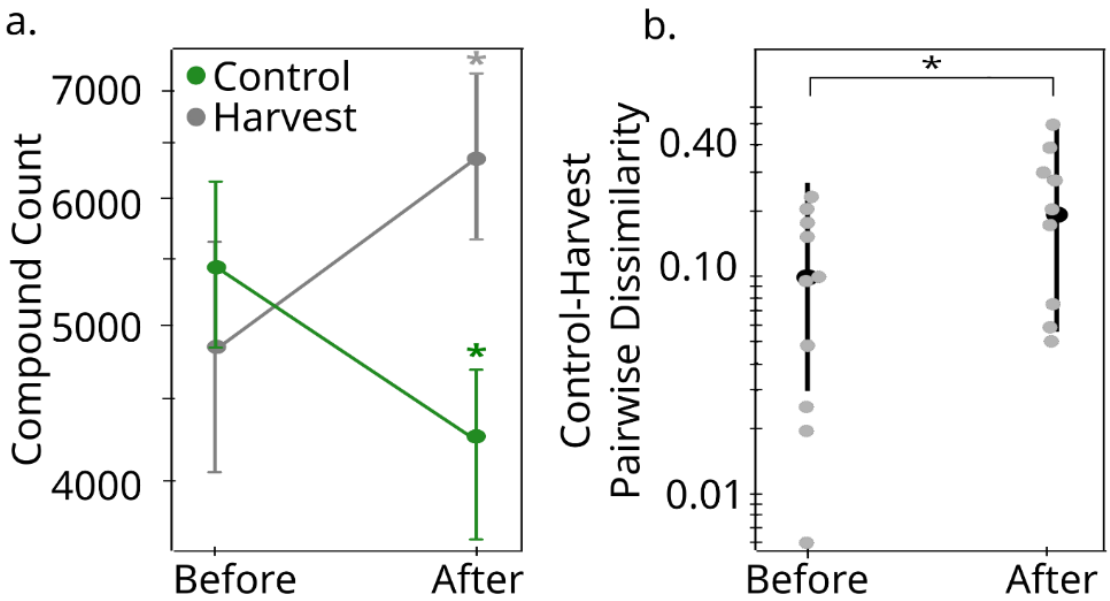


Fig. 3.3 Forest harvest increases molecular diversity of DOM in headwater streams.a. Estimated mean (± 95% CI) compound counts in streams draining control and harvested forests before and after logging. The green and grey asterisks indicate statistically significant differences ($p<0.05$) between control and harvest after logging, and before and after forest logging in both control and harvest sites (Table B4). **b**. Estimated mean (± 95% CI) pairwise dissimilarity between paired control and harvest sites both before and after logging, with a logarithmic y-axis(Table B5). Grey points are individual observations at each time point based on the Bray–Curtis dissimilarity (Fig. B3). The asterisk indicates a statistically significant difference ($p<0.05$) between the mean pairwise dissimilarity before and after logging using a mixed effects model that considered temporal autocorrelation.

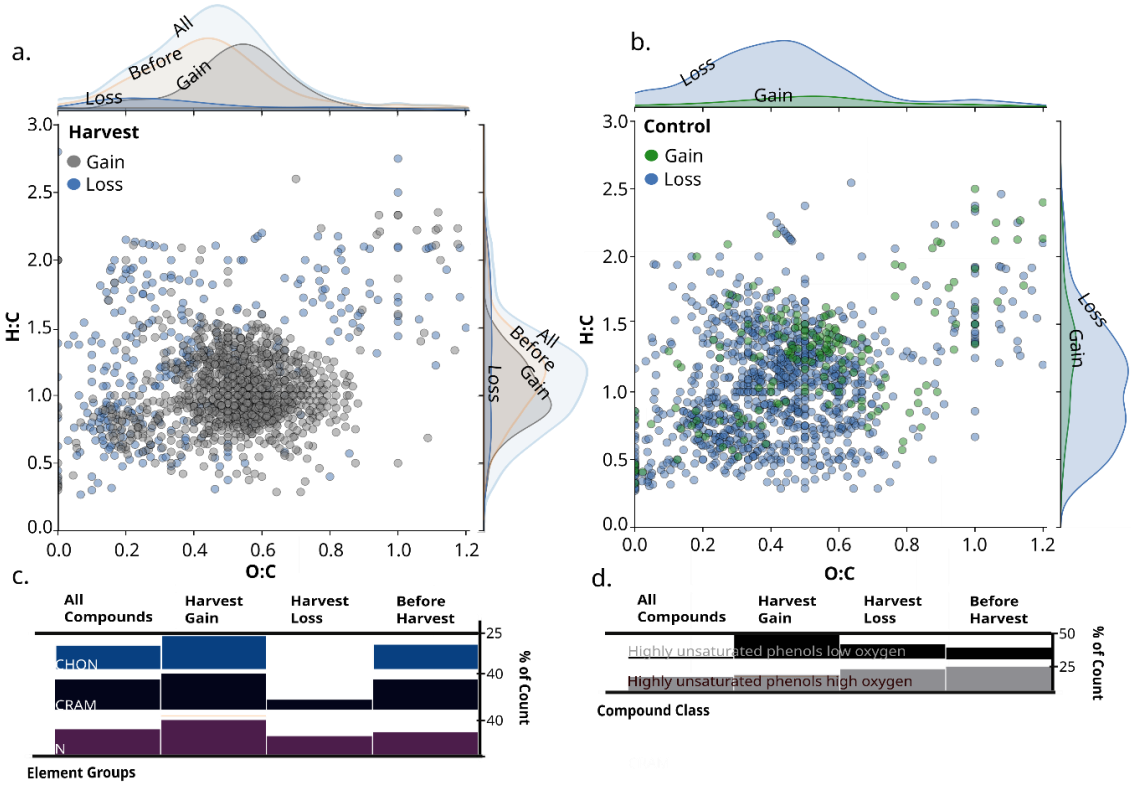


Fig. 3.4 Harvest increases the chemodiversity of stream DOM by introducing novel compounds reflective of fresh plant material and disturbed soil. Elemental ratios of molecular formulae detected in **a.** harvest and **b.** control sites. Probability densities are given along each axis for compounds gained after harvest only in harvest sites (grey, n=1035) and gained after the harvest period only in control sites (green, n=220), for compounds lost only in harvest (blue, n=320) and control (blue, n=962) sites, and all compounds present in the harvest site before logging (n=4927, **a.** only). We also plotted the percentage of all compounds (n=7444) that contained **c.** different elements (Table B7) or classified into **d.** different compound classes (Table B8). Ticks are aligned to the highest bar in each row. The base of the bars equals 0%.

3.3.4 The terrestrial source of increasing stream chemodiversity

To test further if the increased chemodiversity originated from deeper soils after logging, we tracked DOM from soils into streams monthly during the experiment. We did so by installing lysimeters at 5, 15, 30, and 60 cm depth at shoulder, back, toe, and foot hillslope positions at all four catchments (more details in ref[285]). At each time point, we computed the similarity between all soil samples (n=16) and the stream

within a catchment (16 pairs × 4 catchments). We found that the similarity in the molecular composition of DOM between catchment soils and streams persisted in harvest sites compared to controls after logging, suggesting higher hydrological connectivity (Fig. 3.5a). The connectivity between streams and soils in the harvested sites likely arose from intermediate (15 cm) and deep (60 cm) soils remaining connected to streams, given the similarity in DOM composition at these depths in the harvest sites compared to the controls post-logging (Fig. 3.5b). This observation was consistent with the observed increase in nitrogen-rich and CRAM compounds from sub-surface soils in post-harvest conditions (Fig. 3.4). By contrast, streams in unharvested sites declined in their similarity to catchment soils over this same time period from a mean of 60% (95% CI: 58-63%) similarity before logging to 53% (95% CI: 51-56%) afterwards (Fig 3.5a). The loss in similarity between soils and the stream is consistent with the decline in compound count being attributed to fewer source areas that are hydrologically connected to the control streams (Fig 3.3a). We observed an increase in compound diversity with harvest (Fig. 3.3a), and infer that, based on stream compound community similarity to soils (Fig. 3.5b), hydrological connectivity is maintained in harvest sites relative to the controls. The increase in compound diversity with increased hydrological connectivity is the opposite pattern to that observed in rivers, which tended towards less diversity during high-flow periods as more DOM was shunted through the fluvial network[59,286].

An alternative explanation for the maintained soil-stream similarity in the harvest sites relative to the controls is that tire or track forces from harvest machinery caused soil displacement and rut formation[287]. Rutting may partly be responsible for the signal from deeper soils, as ruts on slopes are preferential routes for runoff, which become deeper because of erosion[288]. By contrast, the residue from logging slash left at the soil surface may have contributed to the increased representation of polyphenols.

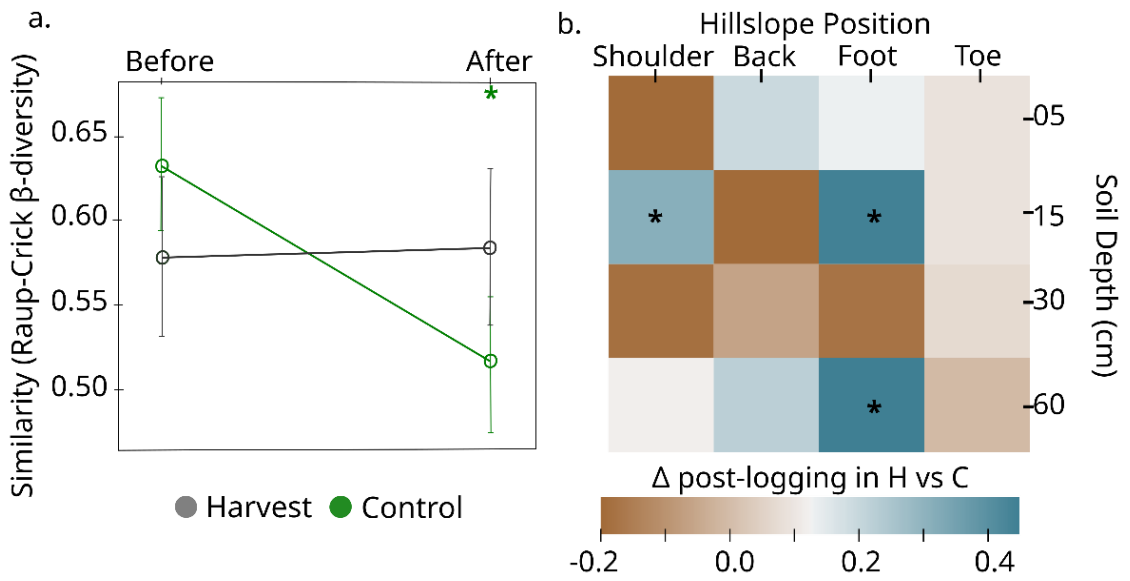


Fig. 3.5 Harvest maintains similarity in DOM composition between stream and soil waters. We calculated the similarity between the molecular composition of DOM in streams and porewater at each soil position (n=4 depths × 4 hillslopes) in each catchment (n=4), accounting for differences solely because of the number of compounds in a sample. **a.** We estimated mean (± 95% CI) similarity across all time points both before (n=5) and after (n=3) logging. * indicates a statistically significant difference (p<0.05) between before and after forest harvest (Table B9). **b**. Changes in molecular composition of DOM after logging in harvest versus the control sites for each depth-by-hillslope combination averaged across catchments. Positive values indicate greater differences in DOM composition after logging in harvest (H) than in control (C) treatments, whereas negative values indicate the reverse. * statistically significant difference in soil-stream similarity between the before and after period between treatments (p<0.05, Table B10).

3.4 Discussion

3.4.1 Losing forest soil carbon to the aquatic network

By disrupting soil, forest harvest breaches spatial barriers that would otherwise isolate soil organic matter compounds from aquatic ecosystems. In deeper soil horizons, SOC is typically stabilised via organo-mineral complexation in addition to physical disconnection from decomposers and enzymes[289,290]. Logging activity removes surface soils, exposing deeper layers (as with ruts), or brings them to the surface. The latter outcome results in the breakdown of organo-mineral complexes

through the loss of physical protections and the export of deep soil DOM that may be chemically biolabile in aquatic ecosystems[291]. Here we also found that DOM from harvested catchments had unique molecular formulae relative to unharvested catchments and these formulae were absent in the streams prior to harvest. This pulse of novel molecules to aquatic ecosystems could have unexpected consequences on the ecology of molecules, analogous to the introduction of novel species to ecosystems[292]. These novel compounds from typically unconnected soil environments may impact stream microbial communities differently than in the compounds' native soil "range", potentially reshuffling the network of relationships amongst DOM compounds and stream microorganisms. Uncovering the relationship between novel (and bioactive) soil-derived additions from deep soils into stream microbial communities is a necessary next step in quantifying carbon losses to the atmosphere resulting from forest harvest.

Although we did not test the direct consequence of additions of dead organic matter sitting on soil surface after logging operations (i.e., harvest slash), long-term experiments suggest that much of it may be rapidly leached as DOM in temperate forests[293]. These inputs of leaf litter have also been found, unexpectedly, to prime the microbial community to breakdown additional soil organic matter (SOM) and change its composition[294]. Leaching of bioavailable plant material through hydrological flow after forest harvest may have similarly unexpected effects on metabolism in water and sediments. For example, Emilson et al[295] found that fewer dissolved phenolics leaching from terrestrial organic matter increased methane production in aquatic sediments. If inputs of harvest slash reduce SOM through microbial priming, there may also be accompanying changes in the composition of leached DOM that impact aquatic metabolism. Tracking the fate of DOM derived from slash is therefore key to understand the efficacy of natural climate change solutions such as forest harvest followed by long-term carbon storage in wood biomass.

3.4.2 Considerations for forest management practices and policies

Our finding that logging impacts the similarity between soil and stream DOM at a molecular level highlights the need for more integrated land-water management. There is a history of managing forests for the protection of water resources in many

countries through best management practices (BMPs) primarily targeting the reduction of soil disturbance, the retention of riparian buffer strips, and the reduction of road-related runoff[296]. While these management practices are generally effective at reducing sedimentation[297], we have shown demonstrable effects on the molecular composition of DOM, with potential implications for the fate of soil-derived carbon. Surprisingly, we found that hydrological and biogeochemical alterations occurred within the soil-water interface of logged forests even with low-impact, selection cut techniques in tolerant hardwood stands with BMPs implemented to protect downstream waters. Therefore, our results suggest there is a need to revisit sustainable forest management practices and policies to consider leaching of soil carbon into aquatic networks. Without these considerations, natural climate change solutions for carbon sequestration that use forests have the potential to be poorly applied and offset in downstream waters.

3.5 Materials and methods

3.5.1 Site description and selection

The four experimental hardwood-dominated catchments were located northeast of Lake Superior, Ontario, Canada (Fig. 3.1) on the northern edge of the Great Lakes-St Lawrence Forest Region[298] the second largest forest region in Canada. We selected the Batchawana Watershed due to the presence of the Turkey Lakes Watershed (TLW) experimental study[299], which was established in 1979 and is one of the longest running watershed-based ecosystem studies in Canada[300] The extensive scientific and support infrastructure at TLW hosts a comprehensive environmental data record which we utilized to estimate C fluxes and understand likely baseline conditions for our sites. The forests of the regions are composed of uneven-aged hardwood forests with primarily podzolic (spodosols) soils with well-developed forest floor horizons[208]. We focus this study on hardwood-dominated forests in the hemi-boreal region because of the expected shift towards hardwood species in temperate forests and the associated increased pressure on hardwood forests[301], and because this forest region is expected to undergo particularly rapid climate shifts[209]. The most recent operational logging occurred in the 1950s when the area was selectively logged for high-quality yellow birch (*Betula alleghaniensis*),

sugar maple (*Acer saccharum*), white spruce (*Picea glauca*), and white pine (*Pinus strobus*)[302].

To select experimental catchments, the regional commercial forestry company (Boniferro Mill Works Inc., Sault Ste. Marie, Ontario) provided historical and future harvest plans. Within the future harvest region, we delineated catchments of headwater streams. Catchment boundaries were delineated using WhiteboxTools[302], a python scripting API for geospatial analysis, along with 30m digital elevation models from the Government of Ontario and existing hydrological data from the Ontario Integrated Hydrology (OIH) dataset. Flow accumulation and flow pointer grids were generated using the D8 algorithm, where flow entering each cell is routed to only a single downstream neighbour[301]. Streams were extracted using the generated flow accumulation map and a threshold value of 0.128km^2. This threshold value was selected by stepping through values and visually validating the generated network against satellite images of streams, catchments generated in smaller areas from 5m LiDAR-derived digital elevation models (DEMs) available for the TLW, and the stream network shape from the OIH dataset.

We then matched the harvest catchments to control catchments by maximising their similarity in a) topography, b) forest composition, and c) ecosite composition. Topography results were generated for each grid cell in the input DEM and included the slope gradient (i.e., steepness in degrees), aspect, sediment transport index, topographic wetness index and Pennock landform classification[266]. For each topographic variable we calculated median, maximum, minimum, and standard deviation values which were then assigned to the derived catchments. Forest composition was generated using the Ontario Forest Resources Inventory (FRI). This inventory is based on digital aerial photo interpretation and field surveys, with a recent incorporation of LiDAR data. From this layer we extracted polygons of overstory and understory species composition (and averaged the percent composition when both were available for a single site), tree species identified in the stand (or the uppermost canopy if the stand contained two or more distinct layers), and the percent cover each tree species occupies within the canopy. Additionally, we utilised the primary ecosite attribute from the FRI database. Ecosite is defined as an ecological unit comprised of relatively uniform geology, parent material, soils, topography, and hydrology and consists of related vegetation conditions. An

intersection function was then used to calculate the percentage of area contributed by each tree species and ecosite to the catchment. We excluded possible control catchments based on a 300m distance to passable road (a semi-arbitrary threshold selected based on logistical concerns). Road data were downloaded from the Ministry of Natural Resources. Catchments within 50m of the DEM edge were also removed to avoid edge effects. These steps left n = 2177 catchments for further selection.

Control catchments were then selected using a divisive (top-down) hierarchical clustering analysis. This procedure is defined by a stepwise algorithm which starts with one cluster of all observations, and splits clusters into subgroups with each successive step based on within-group similarity and inter-group distance defined by the Ward's method linkage and Euclidean distance, respectively. Permutations continued until there was only a single catchment remaining in the same branch/cluster as the catchments selected for harvest.

Within each of the four experimental catchments, hillslopes were further partitioned into four topographic features based on morphological features of surface DEMs according to Conacher and Dalrymple 1977[303] and the Height Above the Nearest Drainage (HAND) terrain model[255]. The HAND model correlates with the depth of the water table, providing a spatial representation of soil water environments (k= 42 clusters)[304]. The four features were shoulder, backslope, footslope and toeslope.

3.5.2 Logging

All harvest operations were performed according to the Ontario Stands and Site Guide for tolerant hardwood selection cuts[212]. Selection cut aims to remove, on average, 30% of the total basal area of tree stems in a stand but never more than 33%. Trees were felled by an Avery Tigercat LX380D feller buncher. Limbs were removed on site, and the tree-length stems were forwarded to roadside landings by rubber-tired skidders with tire chains. Operators were careful to avoid driving machinery into or across stream beds with adherence to the Ontario Stands and Site Guide[305], including leaving 30m buffers on all mapped or obvious watercourses and avoidance of sensitive wet soils following Stands and Site Guide[89].

3.5.3 Stream water samples

Stream water samples were collected monthly during ice-free seasons beginning in September 2019. Surface water (500mL) was grab-sampled at the bottom of each hillslope at the channel head. DOM quantification and molecular characterization was performed on water filtered through 0.45 µm glass fibre syringe filters (Kinesis Inc, USA) into pre-combusted (4 h, 400 °C) amber vials and acidified to pH 2 with HCl within 24 hours. Samples were stored at 4 °C in the dark until analysis within four months. DOC was quantified using a Shimadzu TOC-L (Shimadzu, Japan) with concentrations determined via an L-arginine standard curve. For molecular characterization, we extracted 10 mL of each sample onto 100mg solid-phase extraction columns (Bond Elut PPL, Agilent) and eluted the sample with 3 mL methanol (ULC grade) using the methods described in Dittmar et al[306]. We also characterised DOM with optical analyses of filtered and refrigerated water, described above. Three-dimensional fluorescence scans were performed using a Cary Eclipse (Varian Instruments, USA), and absorbance measured with a Cary 60 UV–Vis (Agilent Technologies, USA). Fluorescence scans were run at 5 nm excitation steps from 250 to 450 nm, and emissions were read at 2 nm steps from 300 to 600 nm. Spectral corrections (instrument and inner-filter), and calculation of the humification index (HIX)[120,307] were applied following standard procedures using the staRdom R package[305]. Finally, major ions, nutrients, and metal concentrations were measured from a subset of the sampled water at the Great Lakes Forestry Centre, Sault Ste. Marie, Ontario, according to methods outlined in ref[257].

3.5.4 Soil water collection

At all hillslope positions, at least 60 mL of soil water was sampled at 15, 30 and 60 cm depths with tension lysimeters that consisted of 60-mm-long round bottom necked porous cups with an outer diameter of 48 mm and an effective pore size of 1.3 µm (model 0653X01-B02M2, Soilmoisture Equipment Corp). The sampling bottles were evacuated to a negative pressure of 50 kPa with a hand pump so suction pressure was ca. 50 mbar above the actual soil water tension. At 5 cm depths, lysimeters could not be securely installed. Therefore, we sampled pore water with micro-tensiometers designed to extract fluids non-destructively from soils using a vacuum[306] through a 0.6 µm ceramic cup (Rhizon CSS samplers, Rhizosphere

Research Products, Netherlands). All the water samplers were installed in triplicate and water pooled at each of the four depth and position combinations to retrieve sufficient volume for analysis. The hillside design of 16 soil samples and 1 stream sample was replicated across the 4 catchments for a total number of 68 samples per sampling date. Soil water samples were subsequently analysed for DOM concentration, fluorescence and molecular characterisation.

3.5.5 FT-ICR-MS analysis

Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR MS) analysis was performed on a 15 T Solarix (Bruker Daltonics, USA). The system was equipped with an electrospray ionization source (ESI, Bruker Apollo II) applied in negative ionization mode. Samples were diluted to yield a concentration of ~5 ppm in ultrapure water and methanol 50:50 (vol/vol). This dilution was filtered through pre-cleaned and rinsed 0.2 µm polycarbonate syringe filters before analysis performed in random order. Electrospray ionization in negative mode (Bruker Apollo II) was done at 200 °C and the capillary voltage was set to 4.5 kV. The sample was injected at a flow rate of 120 µL h^{-1}, the accumulation time was set to 0.05 s, and 200 scans were co-added for each spectrum in a mass range of 92–2000 Da. Each spectrum was internally calibrated with lists of known masses. Mass spectra were exported from the Bruker Data Analysis software at a signal-to-noise ratio of 0 and molecular formulae were assigned using the ICBM Ocean tool[308.] The method detection limit was set to 3. Junction of mass lists along mass to charge ratios (m/z) was performed via fast join at a tolerance of 0.5 ppm while standard smooth and additional isotope tolerance was 10‰. Singlet peaks occurring only once in the dataset were removed, then molecular formulae were assigned with a tolerance of 0.5 ppm in the range m/z 0–1000 within the limits $C_{1–100}H_{1–100}O_{0–70}N_{0–4}S_{0–2}P_{0–1}$. Before statistical analysis, relative intensities were normalized to the sum of intensities per sample. Furthermore, molecular formulae were only retained in the dataset if they occurred at least three times across all samples.

For each molecular formulae, we calculated 10 traits related to molecular weight, stoichiometry, chemical structure, and oxidation state. These traits were molecular mass, the heteroatom class, double bond equivalents (DBE = number of rings plus double bonds to carbon, DBE[309]), carbon number (C), Standard Gibb's

Free Energy of carbon oxidation (GFE)[309], nominal oxidation state of carbon (NOSC), O:C ratio, H:C ratio, and AI_{mo}[310]. Molecules were classified as CRAM where DBE:C was between 0.30-0.68, DBE:H was between 0.20-0.95 and DBE:O were between 0.77-1.75[311]. Formulae with $AI_{mod} \leq 0.5$, 0.5< to ≤0.66, and >0.66 were defined as highly unsaturated and phenolics, polyphenolic and condensed aromatic respectively. Formulae with $1.5 \leq$ H:C ≤ 2.0, O:C ≤ 0.9 and N = 0 were defined as aliphatic[312]. Formulae with O/C > 0.9 were defined as sugar-like whilst peptide-like were defined as $1.5 \leq$ H/C ≤ 2.0, and N > 0[313].

To compare DOM between each soil sample and stream at each time point, we calculated the Raup-Crick dissimilarity metric β_{RC}[314]. This metric estimates whether pairwise samples are more different in composition than expected by chance, given random draws from the compound pool at a site. For any given pair of soil and stream samples, β_{RC} was calculated by comparing the observed number of shared molecular formulae to the number expected by randomly sampling the same number in each site from the entire compound pool following the protocol described in ref[99]. The probability of sampling a compound was based on its abundance across all soil sites. We repeated this sampling 1000 times to measure the similarity between communities in streams and soils compared to a null expectation.

3.5.6 DOM yield and wood carbon estimates

Monthly DOM yield estimates were estimated for the logged experimental catchments by using monthly precipitation and runoff values (mm $month^{-1}$) in the nearby TLW (Fig. B1), with similar size (ranging from 3.5-43.0 ha, median = 2.14 ha, mean= 6.25 ha) and forest composition as our study catchments. Daily precipitation data were collected at 1.5 km south-east of the TLW boundary according to methods described in ref[315] and averaged by month. Streamflow from the TLW catchments was estimated using stage-discharge relationships at 90° V-notch weirs, with stage measured at 10 minute intervals using water level recorders[316]. Stream stage was converted to mean daily discharge (L s^{-1}) by dividing catchment area and computing the total runoff for each month, as described in ref[317–319]. Approximate annual DOC yields were estimated by multiplying the estimated monthly DOC by the estimated annual water yields.

We estimated the amount of tree carbon removed from each catchment using an estimate of aboveground phytomass for the region is described within ref[313] Phytomass was then converted to carbon using a 50% (w/w) dry wood conversion based on the canopy dominant sugar maple[314].

3.5.7 Statistical analysis

We tested if stream water chemistry varied with logging. We fitted linear models to stream water chemistry (including chemodiversity) and stream water DOC. The time period (before or after logging) and treatment (control or impact, i.e., logging) were fixed effects and we included the interaction between these two variables. We also included the number of days since logging as a predictor as well as the three-way interaction, i.e., before-after × control-impact × time, as described in ref[315]. We accounted for repeat sampling of the same sites through time by including a site-level random effect and estimated a continuous autoregressive correlation structure to account for temporal autocorrelation within year and site. Models were fitted with the lme function in the nlme[235,320] package in R[317]. When the predicted variable was a count (i.e., number of molecular formulae), we modelled the error terms using a Poisson distribution and estimated statistical models using the R package glmmTMB[321]. The Poisson error structure was not possible to include with nlme, but, concurrently, we did not use glmmTMB for all statistical models since it does not allow for a continuous correlation structure with time. In the case of glmmTMB, we instead used ordinal values (order of sample points in time) to model the autoregressive temporal correlation structure as properly modelling the error distribution was more important.

Chapter 4

Forest type and not harvest history best predict molecular-signature-verified PARAFAC components in temperate and boreal headwater streams

4.1 Abstract

Dissolved organic matter (DOM), a mixture of organic carbon molecules, plays a key role in aquatic ecosystems and undergoes molecular composition changes due to human activity on land. It is recognised that natural features such as the amount of forest cover and the extent of wetlands influence the molecular makeup of DOM. However, further research is required to understand how DOM in streams is affected by landscape operations like timber harvesting. This line of investigation can inform when optimal management approaches in forestry should be used to effectively safeguard stream water quality, particularly in terms of minimising changes in DOM molecular composition. Because of their comparably high costs, ultra-high-resolution mass spectrometry (UHR-MS) techniques have yet to be utilised extensively in landscape-level studies. To determine how important anthropogenic vs. natural factors are in driving the molecular composition of DOM, the resolution of molecular-level information at the landscape scale will be necessary. We related DOM molecular composition to fluorescence components derived from a cheaper-to-obtain parallel factor analysis (PARAFAC) model built from over 4500 samples in headwater catchments of the temperate-boreal region of Canada. We discovered that a large portion of the molecular data resolved by UHR-MS, which identifies molecular classes (particularly polyphenols, black carbon, and carboxyl-rich alicyclic compounds) and putative source regions within catchments, was captured by our PARAFAC model. However, we also discovered that there was a significant overlap

and redundancy amongst PARAFAC components. The value of defined PARAFAC models, like the one described here, where specific molecular correlations within each component are found and presented, is highlighted by this result. Additionally, while harvest activity may predict specific PARAFAC components within forest type, soil wetness and tree species composition were the main drivers of the models at the landscape level. Together, our findings support the use of fluorescence spectroscopy and PARAFAC modeling, when verified by UHR-MS, as an effective method for observing how freshwater ecosystems respond to forest change, particularly those driven by climatic and tree species range shifts.

4.2 Introduction

Forest ecosystems are critical carbon (C) reservoirs, storing 40% of the planetary soil C stock[319–321], and more than 3 times the C of the global atmosphere and more than 4.5 times that of other vegetation[16]. However, the ability of forests to maintain or increase this critically important soil C reservoir is affected by land use and forest changes, such as forest harvest[322,323]. Economists predict that the demand for wood products will increase by 19% in the coming decade[324,325], and harvest from plantations and secondary and primary forests will rise to meet this demand[324]. This expected increase in harvest presents a potential problem for water resource management, as forest harvest negatively affects the quantity[64,326–333] and quality[334] of C in adjacent waters by disturbing soils and catchment hydrology. Thus, understanding and monitoring how forest harvest alters the forest-water C nexus are fundamental to protecting some of our most valuable and vital resources, such as water supplies.

Once in water, terrestrial C, in the form of dissolved organic matter (DOM), is much more reactive than it is in soils and is, therefore, more likely to return to the atmosphere[335]. Given the urgency to reduce C emissions, protect current C stores, and remove C from the atmosphere, understanding the factors that control the flux of C as DOM from land to water is essential for combating climate change. In particular, there is a need to understand potential management interventions on land that would prevent C loss or alter the fate of forest C once in waters. One set of management interventions that is of specific interest is the application of best harvest management practices across different forest types and under varying harvest intensities.

Measuring DOM and its structural properties can reveal the consequences of harvest on water quality and C cycling, especially with different management practices. DOM is a complex mixture of different compounds that can be used to infer information about the mixture's source and about changes in source area (for example, after harvest, which re-routes hydrological flows[249,336]). Ultra-high-resolution mass spectrometry (UHR-MS) resolves DOM into individual molecular formulae, making it a vital tool in elucidating changes to DOM sources. However, UHR-MS is relatively expensive, inhibiting large-scale investigations across environmental gradients, which are needed to test harvest impacts across forest ecosystems.

On the other hand, fluorescence spectroscopy offers a cost-effective alternative, particularly when the fluorescence signal is decomposed into its underlying components with multivariate analysis, such as parallel factor analysis (PARAFAC)[337]. PARAFAC modeling breaks down measured excitation-emission matrices (EEMs) into individual fluorescent components[329]. These fluorescent components are commonly interpreted to represent different source materials[338]. Leveraging and developing this tool and making it available to practitioners (thereby reducing challenges due to PARAFAC complexity) involved in monitoring land use and forest change could be an effective way to promote progress on critical land and water-use management questions. However, PARAFAC is yet to reach its full potential since studies including high-resolution molecular interpretation of fluorophore components are ecosystem or site-specific[338], and the ability of this approach to detect forest harvest effects on streams across many ecosystem types is not fully established.

Here we investigate the application of Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS)-verified fluorescence-based DOM characteristics to identify key molecular classes for inferring terrestrial sources across forested watersheds with varying harvest intensities and management regimes. We focused our study on the Canadian eastern forest boreal transition, which contains a rich harvest history with partial representation of both boreal and temperate forests[339]. We first collected 4589 water samples from 268 sites, representing a range of dominant tree types (by percent deciduous or coniferous) and climate zones[329]. We included samples primarily from small headwater

catchments (streams, soil pore water, and throughfall). The C in headwaters is almost entirely found in terrestrial DOM previously altered along soil and groundwater pathways[340], which makes identifying the effects of harvest in small headwaters straightforward. However, to representatively sample the region, we also included samples from larger headwater catchments, sampling with a mean ± std of 18.2 ± 40 km^2 (range: 0.3-233.1 km) to build our PARAFAC model. We validated the ability of the model to decompose key DOM properties by identifying associations with molecular compounds from a subset of the samples that yielded straightforward results (n=64 from hillslope soil pore waters and their streams) using FT-ICR-MS. Next, we explored whether PARAFAC components are predicted by harvest history in catchments with known harvest management history (n=187). Specifically, we use random forest models to test the relative importance of forest harvest relative to other landscape factors known to influence DOM molecular composition. Our findings indicate that harvest was not a significant predictor of PARAFAC components at the landscape level, indicating that present management practises are successful in limiting harvest impacts below the natural variability caused by tree type and soil moisture. Harvest, however, emerged as a significant predictor when models of PARAFAC components were run within subsets of catchments defined by forest type and wetness. This result suggests that PARAFAC modeling could be a tool for testing and investigating ecozone-specific best forest management practices aimed at minimising changes to stream DOM quality.

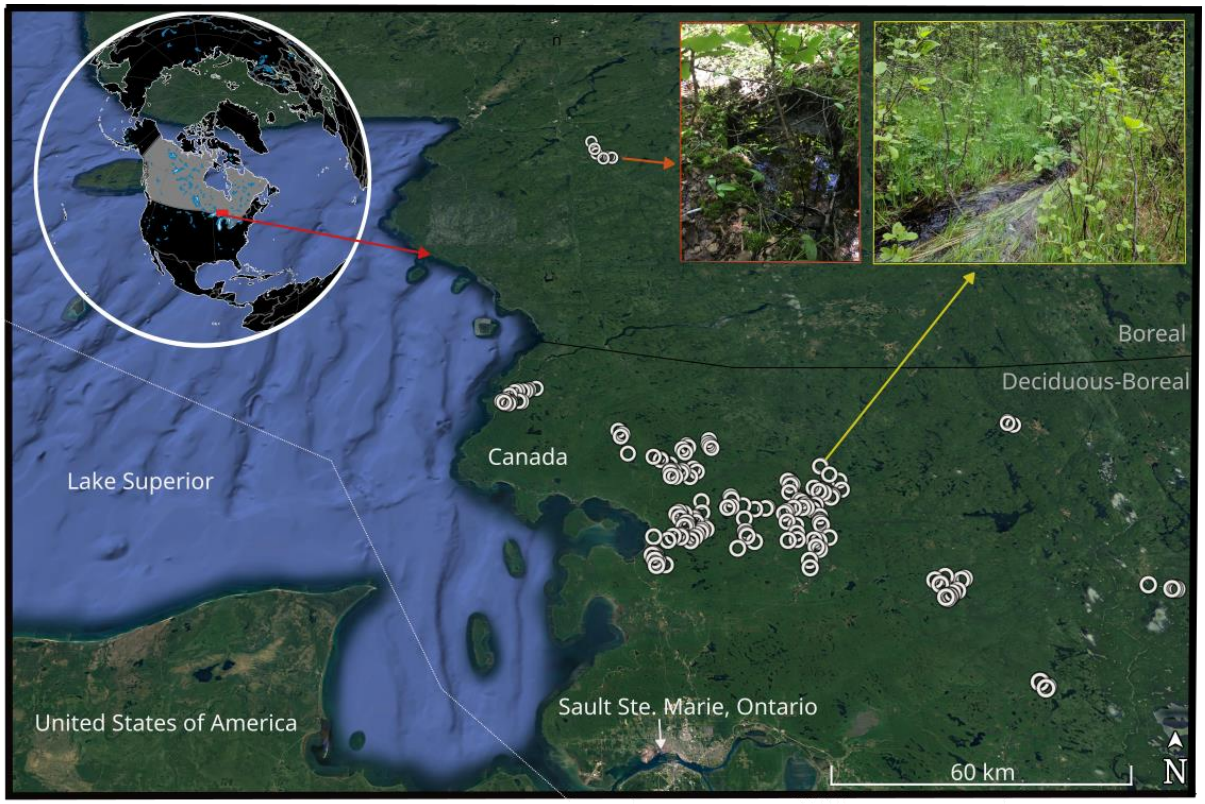


Fig. 4.1 Streams sampled in Ontario, Canada's eastern forest boreal transition. White circles represent sampled streams with corresponding catchment characteristics (n=187), and photographs of two sampled low-order streams are displayed in the upper right corner. All sample sites used to build the PARAFAC model are mapped in Fig. A1. The white line is the national divide. The black line is the approximate division between the boreal and the deciduous-boreal ecozone[341].

4.3 Results

4.3.1 DOM optical characterisation

We captured over half of the variation in typical global DOM concentrations and most (67% variation) within the expected boreal range as well as a representative range for DOM optical properties [342,343–345]. We found that concentrations of DOM in our streams varied by a factor of more than 25, with a mean ± std concentration of 5.3 ± 3.6 mg C L^{-1} (range: 0.8-22.3 mg C L^{-1}). The reported range in boreal stream DOM is 2.3-33 mg C L^{-1} [343,345]. Furthermore, global surveys of rivers typically have a range for DOM between 0.5 and 50 mg C L^{-1} [343–345]. Similarly, the absorbance coefficient at 254 nm (a254: an indicator of coloured DOM) varied over 50-fold, with a mean of

55.3 ± 47.1 (unitless, range 5.2-263.2 m^{-1}), similar to the 9-250m^{-1} range reported by a similar study of streams in the Canadian boreal zone[343,345]. Together this suggests that we representatively sampled the region.

We developed a PARAFAC model with 4,589 samples from Eastern Canada and successfully resolved seven previously identified fluorescent components, explaining 99.3% of the variation in the EEMs (Table 1, Fig. C.2). Here, we present the components as proportions of overall signal intensity with numbers (C1-C7) corresponding to the order of contribution from largest to smallest (Fig. C.2, Fig. C.3, Table C.1, C.2). All seven identified components were compared with previous studies (n=68 unique matching papers) in the OpenFluor database with an index of PARAFAC component similarity (Tucker Congruence Coefficient) for both excitation and emission x excitation > 0.98[329,333] (Table C.3). Components 1 (C1) and 2 (C2) were previously described as widespread humic-like components of terrestrial plant and soil origin, with C1 considered to be of higher molecular weight. Component 3 (C3) has similarly been described as humic-like but associated with photodegraded intermediate products. Component 4 (C4) was also a humic-like component but considered lower molecular weight and more universally observed in terrestrial and marine environments. Components 6 and 7 (C6, C7) were identified as the common tryptophan and tyrosine protein-like components, respectively. Component (C5) had the fewest matches (n=3 studies, Table C.3), with the closest match suggesting it was an artefact from the fluorometer with no ecological implication[329].

Table 4.1 Summary of observed fluorescence components in the PARAFAC model. The selected references represent the original studies that defined specific fluorescence peaks where applicable (also see ref[343,344]) and are named (Component Name) and described (Traditional Descriptions) based on a summary of the probable peak based on matches in the OpenFluor database (Table C.3). Here we provide new molecular descriptions as summaries of results presented in Table 4.2 and Fig. 4.2. CRAM = Carboxyl-rich alicyclic molecules.

Component	Excitation (ex) and emission (em) maxima (nm)	Component Name	Traditional Descriptions	Molecular Descriptions From This Study
C1	ex 350, em 448	UVC humic-like	High-molecular-weight humic, widespread, but highest in wetlands and forested environments, UV-C humic-like, high molecular weight[345–347]	Correlated with black carbon and polyphenols, but few uniquely correlated compounds with no clear molecular pattern. Redundant with C2 and C3 with little added information.
C2	ex 250, em 436	Humic-like	Terrestrial humic-like (Table C.3)	Can be interpreted as plant and surficial soil-derived. Signature of polyphenols and black carbon with 53% overlap with C3, but uniquely associated with high O:C ratio highly unsaturated phenols and low O:C polyphenols.
C3	ex 270, em 508	Humic-like	Humic-like intermediate of the photodegradation of humic compounds (Table C.3)	Can be interpreted as plant and surface soil-derived. Signature of polyphenols and black carbon with 53% overlap with C2,

Component	Excitation (ex) and emission (em) maxima (nm)	Component Name	Traditional Descriptions	Molecular Descriptions From This Study
				but uniquely associated with high O:C black carbon and polyphenols. Strong negative correlations with CRAMs.
C4	ex 325(<250), em 402	UVA humic-like	Low molecular weight, common in marine environments associated with biological activity but can be found in wastewater, wetland, and agricultural environments[52,276,346,347]	Can be interpreted as sub-surface soil-derived. Strong and unique correlations with highly unsaturated phenols, specifically CRAMs, and nitrogen-rich compounds. Potentially useful for drinking water utilities. Strong redundance (78%) with C3, but with relationships in opposite directions.
C5	ex 300, em 414	N/A	Humic-like, a possible artifact from the fluorometer with no ecologic implication (Table C.3)	No unique strong correlations. No clear pattern in correlated compounds.

Component	Excitation (ex) and emission (em) maxima (nm)	Component Name	Traditional Descriptions	Molecular Descriptions From This Study
C6	ex 280, em 352	Tryptophan-like	Amino acids, free bound in proteins, fluorescence resembles free tryptophan, may indicate intact proteins or less degraded peptide material[345–347]	Negatively correlated with low H:C and positive correlations with high H:C.
C7	ex 270, em 308	Tyrosine-like	Amino acids, free of bound proteins, fluorescence resembles free tyrosine, may indicate more degraded peptide material[321]	Interpreted as biolabile material. High information redundancy with C2 (79%) and C3 (60%) because of negative correlations with black carbon and polyphenols, but strong and unique correlations with unsaturated aliphatics.

4.3.2 Interpreting the PARAFAC components on a molecular level

We found that PARAFAC fluorescence techniques can offer information about almost half of the molecules measured by FT-ICR-MS. A total of 2737 of the 7412 (40%) molecular formulas identified by FT-ICR-MS were correlated uniquely (positively or negatively) with one component in the PARAFAC model with a minimum Spearman's rank correlation (ρ) > |0.16| at the 99% confidence limit (Table C.4, Fig. C.4). N-containing formulae accounted for 29% of the total formulae, while S and P contained 9 and 7% of the remaining non-CHO content (Table C.4). This finding and the fraction of correlating formulae tracks closely with previous reports of

EEMs-molecular formulae relationships[321] and suggests a fairly consistent fraction of the DOM pool fluoresces in northern temperate and boreal inland waters.

Further, we found that the molecular information captured by PARAFAC components can resolve specific molecular compound classes of DOM (Table 4.2). To focus our interpretation of molecular associations within the PARAFAC model, we applied a threshold for the correlation strength ($\rho = |0.45|$, $p<0.001$). The threshold of correlation strength between PARAFAC components and molecular formulae was selected based on a plateau in the ratio of uniquely correlated compounds versus all compounds (Fig. C.5). We found that, on average, a mean ± std of 18.7 ± 23.1% and as many as 62.3% of formulae measured within a given compound class were strongly correlated to a component (Table 2). The compound groupings with the greatest proportion of strong correlates (compared to the total number detected) were black carbon (62% of total black carbon count) and polyphenols (40% of total polyphenol count), indicating that the PARAFAC components identified here could be especially effective indicators of these compound groups (Table 2). This finding is consistent with others[348] that identified a component (Ex:Em maxima: <275(295):420) corresponding to classical “Peak A” (see ref[345,346]) as a strong correlate of black carbon. In our model, this peak is decomposed into C2 (32.5% correlated in compound class) and C3 (28.0 % correlated in compound class, Table 2).

We found that PARAFAC components are strongly associated with specific compound classes resolved by FT-ICR-MS. Of the FT-ICR-MS-resolved compounds, highly unsaturated phenols were the most abundant class (n=5102 formulas out of a possible 7412). Formulas contained in this grouping cover a large range of biogeochemical qualities. For example, carboxylic-rich alicyclic molecules (CRAM) are non-colored, biorefractory molecules thought to be highly processed by microbes[255], while lignin (aromatic biomarkers of terrestrial plants[331]) also falls within this class. To distinguish the likely source, we differentiated between microbially derived CRAM (n= 3414 total formulae) and non-CRAM unsaturated phenols (n=1688 total formulae) by stoichiometric definitions outlined in ref[103]. We found that component C4 was strongly associated with the microbially derived CRAMs and correlated strongly with nitrogen-containing formulae (34.6 and 57.4% of correlated

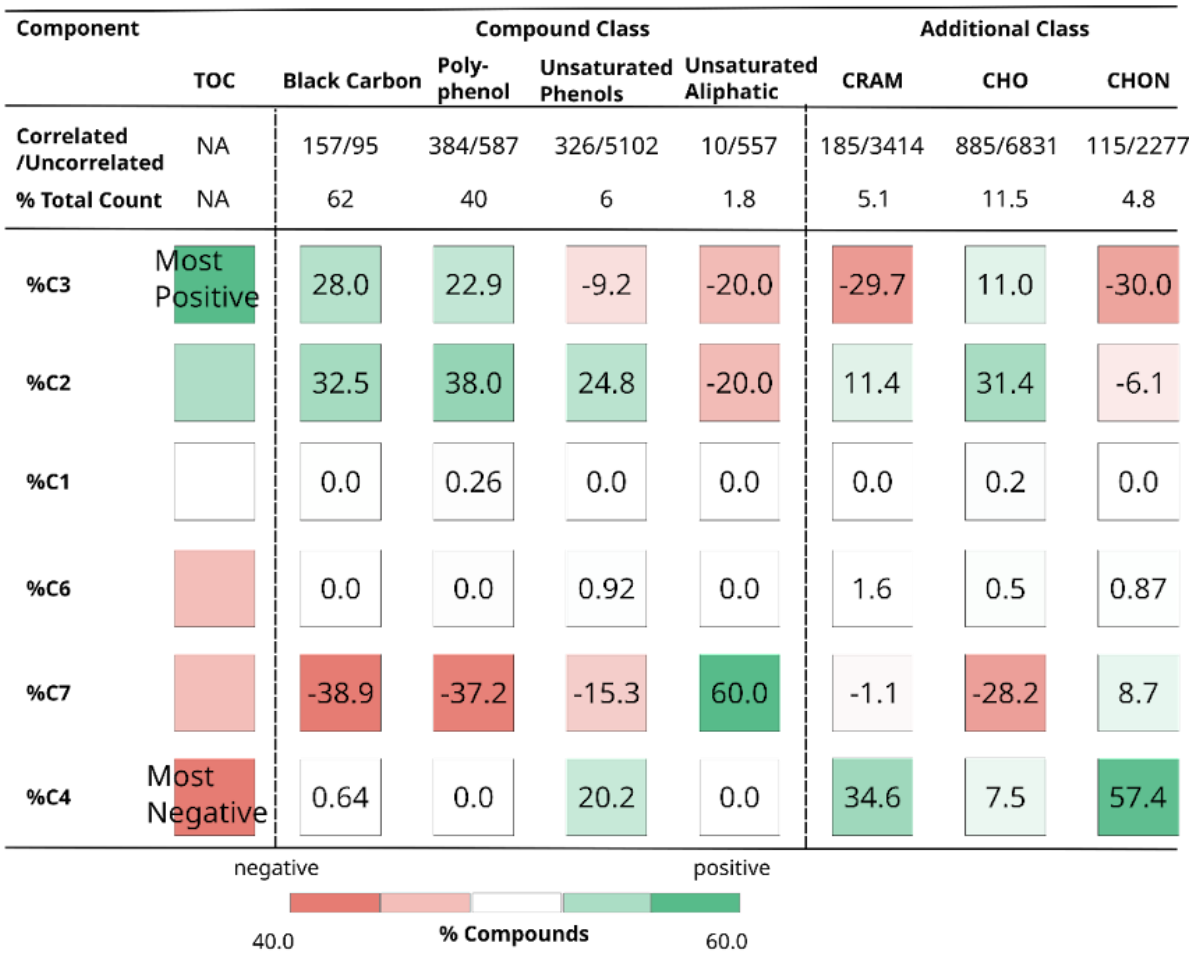
compounds within the compound group, respectively). A recent study of forest harvest impacts on molecules in streams showed an increase in the number of polyphenols and CRAMs following harvest disturbance[349], pointing to the potential utility of C2, C3, and C4 in not only detecting harvest effects but, more importantly, acting as indicators of best-management practice efficacy. Furthermore, given the terrigenous nature of the DOM in the samples, as evidenced by the order of terrestrial components C1, C2, and C3 (numbered based on overall relative intensity), most formulae are likely of plant or soil origin. In this context, signals from C2 and C3 may be interpreted as plant-derived humic material sources from leaf degradation or runoff from surface soil layers. In contrast, deeper soil material and groundwater inputs are likely more represented by C4 because deeper soils are more microbially processed[349]. Finally, we found that C7 strongly correlated with unsaturated aliphatics (Table 2). When present, sugars and peptides also indicated that this component was a unique and strong descriptor of biolabile material (Table 2). Together, the strong association between PARAFAC components and FT-ICR-MS-resolved compound classes points to the utility of components in the PARAFAC model as indicators of compound groupings related to origin within a landscape.

Though useful for pointing to specific compound groupings, the molecular information captured by the PARAFAC components was largely redundant. The range of individual formulas correlated strongly with individual PARAFAC components was between 0 (C5) and 328 (C2) (Table C.4). The molecular formulae that did not correlate strongly with any component (n = 6763) were enriched in peptides and carbohydrates, unsaturated aliphatics, and highly unsaturated phenols (15, 7, and 2 times the amount of the correlated pool, respectively). This exclusion in the correlated pool is unsurprising, given that these compounds are agnostic to fluorescence[350,351]. Of the components, C3 and C4 shared the greatest number of correlated formulas (78%, with 3509 intersecting and 4497 correlated compounds in the union of the set, Table C.5). Still, the percentage of overlap in correlated compounds was greatest between C2 and C7 (3348 intersection of 4244, i.e., 79%). Amongst the most strongly correlated compounds, C2 and C7 also shared the largest percentage of overlapping formulas (74%), but the direction of the correlation with the PARAFAC component was opposite (positive in one, negative in the other). This intersection where the same compounds are correlated with two or more

PARAFAC components suggests information redundancy amongst the fluorescence groupings; it follows then that caution should be applied in interpreting DOM optical signatures alone without accompanying molecular data, as we have provided here

Table 4.2 PARAFAC components are strongly associated with specific FT-ICR-MS-detected compound classes.

Numbers in the cells correspond to the percentage of compounds in a given class correlated with a PARAFAC component at $\rho > |0.45|$ (Tables C.9-C.12). The number of compounds in a group correlated with a given compound vs. the number of compounds that are not correlated to a component is reported both as a ratio and percentage in the top two rows (Table C.8). Correlations are between the components' relative signal and the relative intensity of each molecular formula in a compound class. PARAFAC components are rows arranged from top to bottom by the strength of positive correlation with the total concentration of DOM (first column). C5 was excluded since no compounds were correlated to this component.



Despite the redundancy, we were able to resolve further the specific molecular information captured by PARAFAC components by identifying molecules correlated to only a single component (in a single direction). We detected 452 (6%) compounds strongly correlated with at least one or more components, with 237 positive and 342 negative correlations. To assess redundancy, we allowed a single compound to be considered uniquely correlated once in each direction (Fig. 4.2, Table C.7). We found the most distinct differences in unique correlation between components C2, C3, C4, and C7 (Fig. 4.2). For example, C2 and C3 reflected low H: C ratio formulae, primarily black carbon and polyphenols (Fig. 4.2). However, C2 represented a distinct fraction of DOM compared to C3, with a unique contribution of high O:C non-CRAM highly unsaturated phenols (likely lignin-like phytochemicals[352,353]) and low O:C condensed black carbon and polyphenols. That high O:C unsaturated phenols covary with low O:C polyphenols suggest similarities in their sources and sinks, i.e., they may be biogeochemically related or along the same degradation path[257]. The negative association of C3 with CRAMs further differentiated C3 from C2.

Further, although there was considerable correlation overlap between C3 and C4 (Table C.6), C4 was uniquely correlated to CRAMs. CRAMs are generally unreactive and difficult to remove from water through natural or artificial processes, which is why they are often retained in drinking water after treatment[321]. This result points to the potential utility of this PARAFAC component for water treatment facilities operating within temperate and boreal forests in the case that the direct measurement of CRAMs is not possible. Finally, C7, which overlapped with C2 and C3 because of inverse relationships with black carbon and polyphenols, demonstrated unique and strong positive correlations in the region (of a plot of O:C and H:C) associated with unsaturated aliphatics (which may include lipids and proteins), further indicating that C7 could be broadly interpreted as a signal of biolability.

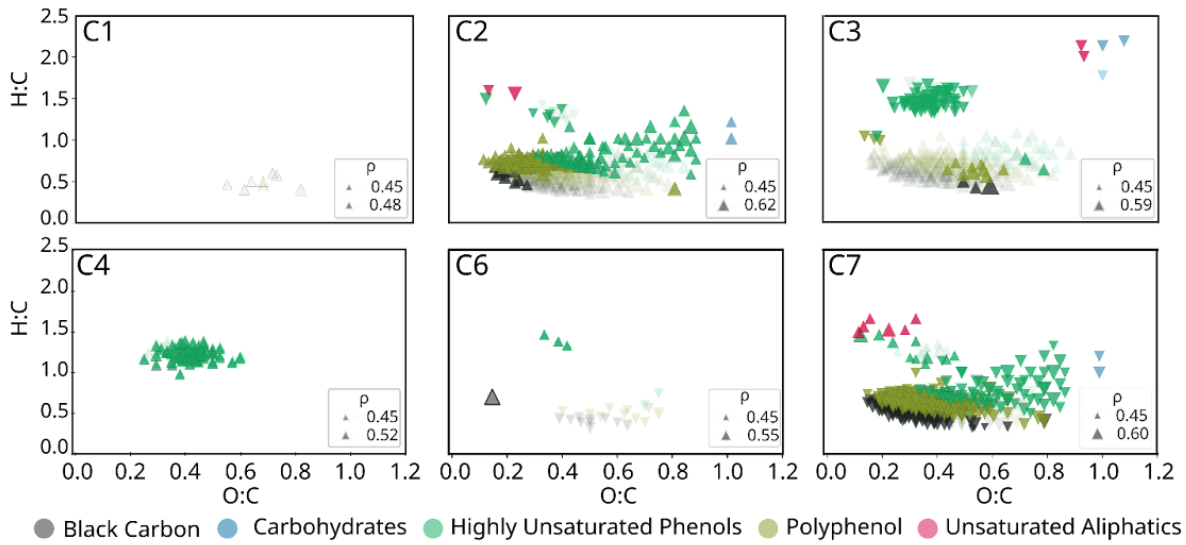


Fig. 4.2 Compounds strongly and uniquely correlated with PARAFAC components (C1-C7). Colours correspond to compound class. Upwards triangles are positive correlations, and downwards are negative. The size of the triangle is proportional to the strength of the correlation for each component. Faded triangles are correlated compounds not unique to a single component. We excluded C5 since no compounds were correlated with this component. Intensities for PARAFAC components were normalised to the sum of component fluorescence intensities within a sample. Mass spectral peak intensities were normalised to the sum intensity of all peaks within a sample. Pairwise Spearman's rank correlations (ρ) were then obtained between normalised PARAFAC components and normalized mass peak intensities.

4.3.3 Predicting PARAFAC components with landscape characteristics

Using the FT-ICR-MS-verified PARAFAC model, we then tested how forest harvest impacts headwater stream DOM, especially compared to other landscape factors known to influence DOM composition (Table 3). We developed random forest models for each component, with catchment characteristics as predictors to disentangle the relative importance of harvest impacts versus natural features (Table 3). We expected that %-forest cover and the presence of wetlands would be strong predictors of the fluorescence signals, as was reported by Kothawala et al.[196] in 17 low-order boreal streams. However, we also included available information on tree species, ecotype (incorporating soil information), and topographic indices (Table C.13). We also integrated harvest impact using two metrics: 1) the cumulative harvest area in a catchment and 2) the cumulative time-adjusted harvest area based on several time-decay functions (see Supplementary Methods). We used time weights to account for the fact that older harvest is less likely to affect stream DOM. Since the true time-decay functions were unknown, we included common decay functions and iterated stepwise through their parameters (rates, cut-offs, see Supplementary Methods). Because all of the catchments were feeding small headwaters and were therefore relatively small (mean size 20 ha), the majority of the harvest in these catchments was located near streams. Hence, we did not spatially weigh harvest regions according to their geographical closeness to streams.

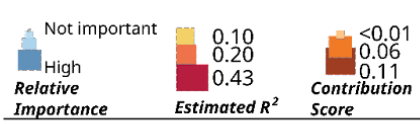
Together, we included 562 input variables (214 landscape, 348 harvest descriptors) to explore how well PARAFAC components in streams could be predicted by natural and harvest features in catchments. We found that of the 562 landscape and harvest variables, 31 were important in predicting PARAFAC components and were broadly related to forest type and the presence of wetlands. Notably, we did not detect an effect of harvest in our random forest approach, despite comprehensively capturing the region's long forest management history with catchments ranging from 0-100% harvest area coverage over the 95-year record (mean: 61.2 ± 35.2%, from 1923). Time-weighting harvest areas also did not improve our ability to detect harvest effects and thus were unimportant in our random forest models (Table 3). This result suggests that at the broad level, landscape controls overrode the effects of land use in these northern transitional forests, including the cumulative effects of harvest. This finding is consistent with results from an

experimental forest area in the temperate forest which also found that differences in DOM composition due to vegetation type and wetland influence were more pronounced than effects due to harvest disturbance[354]. However, the experiment employed only a moderate-intensity harvest treatment (progressive strip cut)[355], whereas the region included in this study included both clear-cut (high-intensity) and selection cut (lower intensity) harvest. We expand the current literature by demonstrating that, on a landscape level, harvest (low to high intensity) is less important than other landscape factors influencing DOM composition in temperate and boreal regions.

The landscape variables important in explaining the PARAFAC components fell into five broad classes, of tree species, landcover, ecotype, and soil wetness. The combined variance explained by all four classes was a mean ± standard deviation of 29 ± 13% (range: 0-43%) of the relative intensity in 5 of the 7 EEMs components that landscape variables could explain. We also predicted DOM concentration due to the relevance of absolute DOM loss from soils to streams in understanding harvest effects on aquatic ecosystems (Table 3). Interestingly, we discovered that two of the PARAFAC components could not be explained by our models (C1 and C6). These were two low-information components in terms of the number of molecules that were correlated (Fig. 4.2). Further investigation will be required to investigate the terrestrial drivers of these components, given this unexpected result.

When considered across the components, the strongest individual predictor was the relative abundance of black spruce (contribution score =0.08, Table 3). The observed sensitivity to tree type has implications for large ecological transition regions such as in the boreal, given that the expected biotic responses to climate change include the poleward migration of species. This poleward migration will happen most rapidly at ecozone edges such as transition zones[255], and especially in disturbed lands within ecozone transitions due to the opportunities for juvenile recruitment of migrating species[355]. Given projections of tree-species range shifts, we may also be able to predict changes to DOM composition in waters. This could prove useful for understanding regions at risk of degradation of both aquatic ecosystem health and drinking water quality.

Table 4.3 Important landscape features in random forest models predicting PARAFAC components and DOM concentration. Important landscape features are grouped by variable type. Rows are the individual important features. Importance was computed as the Gini importance; any variable with a Gini importance score greater than random was considered important. Features that scored lower than random for all components are excluded from the figure. The relative importance of each variable in explaining a PARAFAC component is given by the relative importance score (grey is not important, and dark blue is most important). Columns include the relative signal intensity of PARAFAC components, the DOM concentration, and the contribution score. The fit of the random forest model (below PARAFAC component names) was assessed using the estimated R^2, which measures the amount of variation explained by the model and is represented by boxes ranging from small grey (no variation explained) to large red. The contribution score (final column) was computed as the sum of each row, divided by the number of components (n=8, including the DOM concentration). This metric gives an indication of overall feature importance across all components.



	%C1	%C2	%C3	%C4	%C5	%C6	%C7	DOM	Contribution Score
		0.27	0.11	0.37	0.41		0.16	0.43	
Relative Abundance (Tree Species)									0.11
Black Spruce		0.06	0.09	0.16				0.29	0.08
White Birch							0.05		0.01
White Cedar								0.04	0.01
White Spruce							0.03		<0.01
Larch		0.03							<0.01
Trembling Aspen							0.02		<0.01
Sugar Maple							0.05		0.01
Relative Abundance (Broad Landcover)									0.09
Forest		0.16	0.03	0.16	0.23				0.07
Water					0.13				0.02
Wetland				0.03	0.03				0.01
Relative Abundance (Dry Ecosite Class)									0.10
Dry to Fresh, Aspen-Birch Hardwood							0.03		0.01
Dry to Fresh, Maple Hardwood							0.11		0.01
Dry to Fresh, Red Pine-White Pine Mixwood		0.15	0.14	0.09	0.05				0.05
Dry to Fresh, Spruce-Fir Conifer		0.07							0.01
Very Shallow, Red-Pine-White Pine Conifer							0.04		0.01
Very Shallow, Red Pine-White Pine Mixedwood							0.03		0.01
Relative Abundance (Wet Ecosite Class)									0.02
Moist, Coarse: Spruce-Fir Conifer					0.07				0.01
Moist, Fine: Pine-Black Spruce Conifer		0.02							<0.01
Rich Conifer Swamp		0.03						0.08	0.01
Topographic Indices									0.10
Median Slope				0.03				0.03	0.02
Median Height Above Nearest Drainage							0.03	0.06	0.01
Median Sediment Index				0.07				0.03	0.01
Median Wetness		0.03		0.03			0.03	0.04	0.03
Standard Deviation Sediment Index		0.02						0.03	0.01
Standard Deviation Slope		0.03							0.01
Standard Deviation Wetness		0.03		0.06			0.02		0.01

We subsequently explored if harvest effects within the groups of important landscape variables could be detected and found that the effect of harvest depended on

dominant ecotypes, suggesting the important role in different management regimes (which typically differ by ecozone), soil types, and hydrological connections typical to different ecotypes (Table 4).

Given the importance of tree species (2-29% variance in PARAFAC components explained, Table 3) and soil wetness (2-15%), we split the catchment set according to forest type (defined by black spruce percentage, Table 3), the wetness of soils (topographic wetness index), and if the catchment had been impacted by harvest within the last five years. We selected a five-year time threshold as a recent review suggested that most harvest impacts on solutes in forests with seasonal snow cover occurred within 3-6 years[356]. Within catchments that had been harvested at least once in the previous five years (n=118, or 64% of catchments), we found that variation in the relative proportion of component C4 could be predicted by harvest area (14% of the variation in the relative intensity of C4, and was the 16th most important variable in the model). Notably, C4 was the component most correlated with N-containing molecular formulae and CRAMs (Table 2), consistent with recent observed increases in these compounds in an experimental harvest experiment performed within the eastern boreal transition zone[321]. Soil organic matter at depth is highly enriched in microbial compounds and is more processed by microbes than surface soil organic matter[355], suggesting that, first, our PARAFAC model may be sensitive to signals of soil disturbance in catchments, and, second, observations from forest harvest experiments may scale to the landscape level.

The most robust models of PARAFAC components were those where catchments were split based on wet soils and tree type. In this case, variation in wet black spruce stands that included harvest as a predictor explained 24% of the variation in the relative intensity of C3. Black spruce stands are typically clear-cut[357] and often burn[358]. Therefore, it is unsurprising that the PARAFAC component most strongly correlated with polyphenols, a signal of surface soil inputs[359], and black carbon (a signal of fire[360]) may be predicted by harvest as additional surface soil is mobilised into streams following high-impact clearcutting. Clearcutting is the standard management practice in Ontario black spruce stands[310,361]. The PARAFAC component C4 could, again, be predicted by harvest impacts (17-23%) when excluding dry soils in all (wet) forests as well as in (wet) deciduous-dominant forests where black spruce was excluded (Table 4). Together, this suggests that

components C3 and C4 from our model could be useful PARAFAC components in detecting harvest impacts in boreal and deciduous forests, respectively. Furthermore, these components may help test the efficacy of forest management techniques in protecting stream water quality within boreal and boreal-transition forests.

Resolving forest harvest with PARAFAC components is an inherently challenging task due to the redundancy in PARAFAC components and diverse terrestrial sources introduced due to harvest activities[362]. Models where harvest was an important feature, explained a maximum of 39% of the variance in the relative fluorescence intensity of particular components, indicating the potential of these tools in monitoring forest harvest impact, but large-scale applicability will require further refinement. Despite the limitations, once the catchments are partitioned by the dominant ecotype, harvest effects can better predict fluorescence signal increases.

Table 4.4 Harvest effects predict fluorescence components within subsets of catchments defined by forest type, recent harvest (<5 years), and wetness. Rows are the criterion class used to subset the catchments. Under the criterion names (rows) are the numbers of catchments included in each: “All Forest,” “Black Spruce,” and “No Black Spruce” columns. Columns are defined by forest type, which was determined using Jenks Natural Breaks optimisation (with a single cut-off) on the dominant boreal species (black spruce, *Picea mariana*)[363]. Wetness was defined using Jenks Natural Breaks optimisation (with a single cut-off) on the topographical wetness index. Columns with less than ten catchments were considered to have insufficient sample size for random forest modeling (written as NA). Components that could be explained by the random forest model that included harvest as an important feature are shown for each row x column subset in bold. Harvest impacts include the unadjusted cumulative harvest areas as well as the time-weight adjusted harvest area. “No Components” = none of the relative intensities in PARAFAC components could be explained by the harvest predictors. Components displayed in grey had no clear FT-ICR-MS compound interpretation (Fig. 4.2). The number that follows the component is the proportion of variance explained by the model. Numbers in brackets are the order of importance for the harvest variables (i.e., the n^{th} most important variable) in the model explaining the relative intensity in the PARAFAC component. When more than one value is contained within the brackets, more than one harvest predictor contributed to the variation in relative PARAFAC component signal intensity explained by the model. All

models that explained less than 10% of the overall relative signal intensity variation were excluded.

Criterion	All Forest		Black Spruce		No Black Spruce	
Harvest < 5 years 39/5/34	**No Components**		**NA**		**C4**	0.14 (16)
Wet Soils 144/21/12	**C4**	0.23 (38)	**C3** C6	0.24 (29) 0.16 (33)	**C4**	0.17 (10)

4.4 Discussion

Here, we have presented a PARAFAC model derived from a dataset of over 4589 samples, with associated molecular characterisation, and demonstrated applicability for headwaters in northern boreal and temperate forests (Table 4.2). We suggest that fluorescence spectrometry practitioners operating in temperate or boreal zones fit EEMs to existing PARAFAC models[364], such as the one developed here, where molecular associations are defined. Furthermore, The PARAFAC model presented offered the advantage of unambiguously resolving molecular formulae within the compound classes of black carbon, polyphenols, and CRAMs (Fig. 4.2). Thus, employing this model can be a valuable way of acquiring molecular-level interpretations when direct measurements with FT-ICR-MS or other UHR-MS instruments are not possible. While the universality of our molecular interpretations beyond northern temperate and boreal forests requires investigation, we have created an open and usable model for this global forest region that covers 25% of the world's area and at least a third of the global terrestrial C[365]. Our FT-ICR-MS-verified PARAFAC model also provided insight into ecological and forest management questions. We demonstrated that harvest was relatively unimportant across the eastern boreal transition region compared to forest species composition and wetness in predicting the relative intensity of PARAFAC components. This low

relative importance at the landscape level suggests that harvest impacts on stream DOM chemistry may be relatively minor compared to those expected under future climate changes. For example, the most important predictor across the set of PARAFAC components was the relative proportion of black spruce in a catchment (Table 3), which is expected, by 2060, to be lost within the region due to northward migration[366]. Similarly, wetness (or the lack thereof, Table 4.3) was an important fluorescence predictor, which suggests that climate-driven changes in precipitation patterns[367] will likely exert a strong influence on stream chemistries. Given that harvest impacts were found to be predictors of C4 and C3 PARAFAC components in catchments grouped by harvest history, wetness, and forest type, an intriguing potential use for this model could be to evaluate the effectiveness of best management practices for harvesting forests in these two ecozones. These best management practices could be geared toward minimising the impacts of harvest on the natural DOM composition in headwater streams. However, at the landscape level, current management in the region maintains harvest-driven DOM changes below the natural baseline variation.

The combination of EEM/PARAFAC and FT-ICR MS offers valuable insights into the characterization of dissolved organic molecules, yet despite advancements, inherent limitations suggest vast potential for further refinement and exploration. In a 2014 study, Stubbins et al. examined river water samples from boreal Quebec, Canada. They similarly employed the EEM/PARAFAC methodology in conjunction with FT-ICR MS to define the molecular signatures associated with EEMs. They found that 39% of the molecular formulas identified by FT-ICR-MS were related to PARAFAC components, showing that the coupling of FT-ICR MS and EEM/PARAFAC offered information on not just about the fraction of dissolved organic molecules that fluoresce, but about broader families of biogeochemically related molecules that track one another in the environment. We similarly found that 40% of molecular formulas were related to PARAFAC components which further supports the idea that there exists a core group of aromatic fluorophores that give rise to the fluorescent properties of DOM[329] However, there are still large challenges and opportunities for further advancement and refinement involved in resolving DOM mixtures. First, fluorescence spectroscopy is constrained only to coloured and fluorescence DOM[91]. Second, each molecule’s optical properties and ionisation

efficiency consistency are altered by the environmental matrix in which they are dissolved as well as the unique combination of other organic present in the mixture[106,361,362]. This is especially complex when considering aggregations of loosely associated molecules held in associations with cations, such as iron, fluorescence[363]. Furthermore, FT-ICR-MS is constrained to relatively small-to-mid-sized organic molecules[48]. Thus, the analytical windows of FT-ICR-MS and fluorescence spectroscopy overlap only somewhat and both do not characterise the DOM pool fully[364] highlighting fertile ground for further exploration. As we embrace these limitations, there are promising opportunities for refinement and further understanding of DOM composition.

4.5 Materials and methods

4.5.1 Site selection and DOM sample collection

We collated DOM samples additional boreal, and eastern boreal transition forests analysed using the same instrument and analysis protocol as the 184 collected stream samples, described below. This dataset included 1205 soil pore water samples and 275 precipitation and throughfall samples from 18 catchments in the Turkey Lakes Watershed Experimental Area (TLW). Details on sampling methods can be found in refs[310,365]. In addition, we included 3015 samples from 78 headwater sites across Ontario (including from the TLW)[366], Quebec[367], and New Brunswick[368] (Fig. C.1). Details on sampling methods can be found within the referenced papers. Together, we included 268 sites and 4589 individual samples to build our PARAFAC model. These samples captured seasonable variability with a mean ± std of 382 ± 294 samples taken in each of the 12 months (range: 65–974 samples (Jan and Oct)). Furthermore, these 268 sites covered a range of forest types; the percentage of the deciduous forest was a mean of 72 ± 23 % (range: 3-100%), and catchment sizes with a mean of 7.5 ± 24.6 km^2 (median = 0.4, range: 0.1-233.5 km).

In the summer (June-September) of 2019, we grab-sampled 250 ml of water from 184 headwater streams for subsequent DOM fluorescence measurements. These streams were selected to ensure maximal variation in boreal-transition landscape and catchment characteristics among sites. Variation was maximised first by generating catchments for all streams with an area of less than 10 ha in the

Ontario boreal-transition landscape. Catchments were generated using the hydrological modelling method described in subsequent sub-sections. We then estimated 214 landscape features for each catchment. Next, we generated a dendrogram of agglomerated streams using a hierarchical clustering approach[369] based on catchment similarity across all 214 catchment features. The similarity was determined using the Ward method, which estimates how much the sum of squares increases when two samples are merged[370]. By visually inspecting the sum of squares increase (the merging cost) and the dendrogram, we aimed to sample sites from 14 generated clusters. This method was executed using scikit-learn[371] 1.2.0 in Python (3.8.16). The selection of streams within the 14 clusters was based on logistical optimizations such as distance to the road and distance to streams from other clusters to ensure sampling efficiency. The sampled 184 headwaters and their catchments were used in both the random forest and PARAFAC models.

We utilised FT-ICR-MS data from soil water at 68 different soil positions and depths and within streams (see ref[372] for collection details) to capture terrestrial source area contributions to headwater streams fully. In addition, half of these catchments were also harvested in 2020 (see ref[373]), which ensured that harvest molecular signals could be interpreted within the PARAFAC context.

4.5.2 Contributing land area delineation

Each stream's contributing area (i.e., catchment) was estimated using hydrological models derived from the 30m Provincial Digital Elevation Model (DEM) produced by the Ontario Ministry of Natural Resources and Forestry. All hydrological modeling was performed using *WhiteboxTools* and the advanced geospatial analysis platform and application programming interface (API)[372] from Python v. 3.7. The retrieved DEM was hydrographically corrected by removing spurious depressions and by breaching road-side impoundments[373]. From the corresponding DEM, drainage paths and directions were modeled using the D8 (deterministic eight-node) single-flow direction method[377]. This flow-routing algorithm is the most commonly used that tracks “flow” from each pixel to one of its neighboring eight pixels based on slope gradient[374]. The position of watercourses was modeled based on flow direction and accumulation layers generated from the DEM. First-order streams were defined and

extracted from the flow network using a flow accumulation area of 10 ha as a minimum threshold, and then the stream points were snapped to the extracted watercourse network. The land (watershed) area contributing to these snapped points was calculated using the Unnest Basin tool. This tool can perform a water-shedding operation based on a group of specified pour points, i.e., outlets or target cells, whereby each complete watershed is delineated.

4.5.3 Catchment characteristics

4.5.3.1 Topography and wetness

We predicted the catchment slope (steepness in degrees) for each grid cell from the input DEM[375]. In addition to the slope, aspect, sediment transport index, and Pennock landform classification[379], a measure based on measures of slope gradient were generated.

We predicted soil wetness and wet area using the topographic wetness index (TWI). Originally developed by Bevin and Kirkby[376], this index is defined as $\ln(a/\tan\beta)$, where a is the local upslope area draining through a certain point per unit contour length and $\tan\beta$ is the local slope. The slope gradient, was generated using Horn's 3rd-order finite difference method[377] and the clock-like grid-cell numbering scheme proposed by Gallant and Wilson[375]. The local upslope area was calculated using the D-infinity flow algorithm and produced as the specific contributing area (SCA). The D-infinity algorithm has improved soil wetness mapping for TWI-based wetness maps[378–380]. The SCA is the total upslope catchment plan area draining across a contoured unit (flow width)[381]. High TWI indicates areas of high-water accumulation (low slope), while low TWI indicates steep slopes and low water accumulation. In addition to the TWI, we modeled the elevation of each grid cell above the nearest watercourse cell measured along the downslope flow path. This metric is related to the height above drainage (HAND) described by Renno et al.[303] Multiple flow-initiation thresholds, i.e., the minimum area required to maintain a stream (0.5, 1, 2, 4, 8, 10, and 16 ha[382]), were input into the HAND algorithm to optimise wetness prediction and account for landform differences.

4.5.3.2 Forest cover, wetland cover, forest tree species, and ecosite composition

Forest composition and coverage were generated using the Ontario Forest Resources Inventory (FRI)[383]. This inventory is based on digital aerial photo interpretation and field surveys, with the recent incorporation of LiDAR data. From this layer, we extracted polygons of overstory and understory species composition (and averaged them when both were available for a single site), tree species identified in the stand (or the uppermost canopy if the stand contained two or more distinct layers), and the percent cover each tree species occupied within the canopy. Additionally, we utilized the primary ecosite attribute from the FRI database. An ecosite is a stand-level unit that describes a region of relatively uniform geology, parent material, soils, topography, and hydrology and consists of related vegetation conditions[384]. An intersection function was then used to calculate the % area contributed by each tree species and ecosite to the catchment, respectively. Additionally, we included the percent coverage of several generalised water and land types (for example, wetland and forest) provided by the Ontario Forest Resources Inventory (FRI)[383].

4.5.3.3 Harvest data and land use

Harvest polygons were extracted from the Ontario Forest Resources Inventory (FRI), which maps the timber resources available in Ontario and maintains harvest records (https://www.ontario.ca/page/forest-resources-inventory). We matched harvest polygons (containing information on the date harvested and harvest area) to catchment characteristics. However, we did not account for different harvest types, though even-aged harvesting (clear-cutting and its variants) has been a universally applied practice across the Canadian boreal forest[385]. In contrast, selection cutting is the most common silvicultural practice for hardwood forests (including maple and birch) in Ontario[386].

4.5.3.4 Harvest area time weighting

We utilised simple additive weighing of harvest areas, which refers to a weighted sum of all the harvest areas contained within a catchment during the entire harvest history (1920-2019). The total harvest impact score (IS) for each catchment (IS_c) was calculated by multiplying the proportional area of each harvest polygon (A_i,

proportion of the catchment area) by the temporal impact weight (w_i). We did not know *a priori* what the temporal weight ought to be. Several papers suggest harvest impact can last anywhere from 0-100 years but do not provide decay functions[120,307]. Therefore, we selected four temporal weighting functions commonly used in cumulative impact measures to define w_i (See Supplementary methods for their functional form). We iteratively selected parameters for these functions to feed to the random forest so that the best-predicting weight could be selected without human bias. The first measure was a step decay function, where the weight is either 0 or 1, depending on the cut-off value. We defined the cut-off with an iterative approach, subtracting between 1-99 from 2019 as a cut-off. After the cut-off value, the weight was assigned 0. The second measure was a linear decay, where weights decline linearly until, again, the cut-off is reached; from that point onwards, the weights receive a value of 0. We assigned all possible cut-off values by subtracting 1-99 from 2019 in year timesteps. The slope of the linear decay was set between 0.001 and 0.09, with 50 steps. Using similar iterative approaches, we generated all possible weights for negative exponential and inverse power weighting measures (Supplementary methods) and again used every cut-off value between (1-99) and 50 values for decay speeds. Finally, we sum the weighted area of all harvest polygons within a catchment before running the random forest model:

$$IS_c = \sum A_i w_i \tag{1}$$

4.5.4 DOM measurement and analysis

All DOM samples were filtered through 0.45 µm glass fiber filters. Samples were acidified to pH < 2 with HCl for FT-ICR-MS or left unacidified for PARAFAC and stored at ~4 °C before analysis. For FT-ICR-MS, we extracted 10 mL of each sample onto 100mg solid-phase extraction columns (Bond Elut PPL, Agilent). Next, we eluted the sample with 3 mL methanol (ULC grade) using the methods described in Dittmar et al[107].

4.5.5 Modelling fluorescence components

Fluorescence EEMs (excitation-emission matrices) were generated for all samples using an Agilent Cary Eclipse Fluorescence Spectrophotometer with a 1 cm path-length cuvette. EEMs were generated from excitation and emission intensities (EX:

250–450 nm in 5 nm steps, EM: 300–600 nm in 2 nm steps) adjusted for inner-filter effects with absorbance as measured with an Agilent Cary 60 UV-Vis Spectrophotometer. All EEM sample corrections and PARAFAC modeling were done in R according to the methods outlined in ref[305]. In addition, seven PARAFAC components were validated by a split-half method[324] with a Tucker's Congruency Coefficient of 0.996, explaining 99.3% of the variation in the EEMs.

4.5.6 Molecular formula and compound groups

Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) analysis was performed on a 15 T Solarix (Bruker Daltonics, USA) as described in ref[383,384] Mass spectra were exported from the Bruker Data Analysis software at a signal-to-noise ratio of 0, and molecular formulae were assigned using the ICBM Ocean tool[385]. The method detection limit was set to 3. Junction of mass lists along mass-to-charge ratios (m/z) was performed via fast join at a tolerance of 0.5 ppm while standard smooth and additional isotope tolerance was 10‰. First, singlet peaks occurring only once in the dataset were removed, then molecular formulae were assigned with a tolerance of 0.5 ppm in the range m/z 0–1000 within the limits $C_{1–10}0$ $H_{1–100}O_{0–70}N_{0–4}$ $S_{0–2}P_{0–1}$. All intensities were normalized to the sum of intensities per sample. Furthermore, molecular formulae were only retained in the dataset if they occurred at least three times across all samples. We calculated ten traits related to molecular weight, stoichiometry, chemical structure, and oxidation state for each molecular formula. These traits were molecular mass, the heteroatom class, double bond equivalents (DBE = the number of rings plus double bonds to carbon, DBE[383,385]), carbon number (C), Standard Gibb's Free Energy of carbon oxidation (GFE[383], nominal oxidation state of carbon (NOSC), O:C ratio, H:C ratio, and AI_{mod}[367]. Molecules were classified as CRAM where DBE:C was between 0.30-0.68, DBE:H was between 0.20-0.95, and DBE:O was between 0.77-1.75[97,386]. Formulae with $AI_{mod} \leq 0.5$, 0.5< to ≤0.66, and >0.66 were defined as highly unsaturated and phenolics, polyphenolic and condensed aromatic (black carbon), respectively. Formulae with $1.5 \leq$ H:C ≤ 2.0, O:C ≤ 0.9, and N = 0 were defined as aliphatic[387]. Formulae with O/C > 0.9 were defined as sugar-like, while peptide-like were defined as $1.5 \leq$ H/C ≤ 2.0, and N > 0[388,389].

4.5.7 Statistical analyses

To deal with so-called “small n, large p,” i.e., with far more catchment variables than measurements per catchment, in addition to complex interactions and highly correlated predictor variables[387,388], we trained random forests[389]. This ensemble learning method detects trends in the data of multiple regression trees to predict how relevant landscape predictor variables could explain each PARAFAC component. The random forest algorithm applies bagging (bootstrap aggregating), which selects a random subset of features at each candidate split in the learning process. Because a random subset of variables is selected for each tree, the problem of overfitting is avoided[387,389]. We used 30% of the catchment data as a training set and repeatedly bagged 1000 times. We quantified catchment characteristic importance using variable importance measures, which allow the user to compare predictor variables with respect to their impact in predicting the response[387]. The importance of a feature is computed as the Gini importance (the normalized total reduction of the criterion brought by that feature) and implemented using scikit-learn 1.2.0 in Python[371] (3.8.16). The coefficient of determination for the model was calculated using the subset of test samples, and the remaining training samples in $(1 - \frac{u}{v})$, where u is the residual sum of squares and v is the total sum of squares.

Chapter 5

General discussion

Forests and waters are inextricably linked, and DOM provides invaluable insight into this link. Harvest disturbances are expected to increase in forested landscapes and influence DOM quantity and quality in aquatic environments. This dissertation aimed to advance our understanding of how forest harvest impacts DOM on a molecular basis in northern forest ecosystems.

In **Chapter 1**, I laid out three questions that this thesis aimed to answer:

i. Does the molecular composition of DOM change similarly through soil depth and along hillslopes with different environmental conditions and, thus, suggest a potentially universal degradation process? (**Chapter 2**)
ii. How does harvest disturbance to soils affect the ecology of molecules in streams? (**Chapter 3**)
iii. How important is harvest relative to other landscape factors influencing DOM molecular composition, and do the harvest impacts on molecules in streams scale to the landscape? (**Chapter 4**)

In this last chapter, I discuss how these questions have been answered by the analysis presented here. I also discuss some limitations of my work and provide possible avenues for future research based on my results and emerging priorities and methods.

5.1 Unlocking the complex interactions between microbes, DOM composition, and the environment

In **Chapter 2**, we built a baseline understanding of soil DOM molecular composition along forested hillslopes by demonstrating that microbial processing influences the distribution and, thus, the persistence of organic compounds in soil. We found that the molecular composition of DOM changed similarly through soil depth and along hillslopes, despite environmental differences. This was driven by two processes: the

selective retention of universal compounds and the consistent re-working of non-universal compounds. We found that genes encoding enzymes involved in the breakdown of plant cell walls were especially related to mass shifts of microbial-produced compounds, e.g., carbohydrate-like rather than lignin- or tannin-like.

The work performed in **Chapter 2** is not only a vital step in establishing base conditions needed to assess the effects of harvest (as tested in **Chapter 3**) but also an advance in understanding the role of microorganisms on DOM molecular composition. Prior work has established that microorganisms (with the assistance of UV radiation) are the main engines of the biogeochemical cycle. Yet, it was unclear if they act to diversify DOM molecular composition or to homogenise it[97,390]. Diversification has been suspected previously based on the myriad of organic compounds that microorganisms produce while alive, as well as those released upon death[391]. The production of these compounds should be determined by the composition of the microbial community (as determined by deterministic and stochastic evolutionary and dispersal events) and their enzymatic potential. Given the number of potential substrate molecules and the multitude of microbial organisms[392,393],microenvironments[394,395], and their interactions with one another[396], the number of possible pathways for DOM degradation is unwieldy. The presence of general degradation patterns along the land-ocean continuum that we observed beginning in soil pore waters and along hillslope flow paths provides new evidence that DOM converges towards a universal compound pool as microbial reworking removes source-specific components. This process provides the mechanistic basis for popular heuristic models like the river continuum concept[396] and soil chromatograph[274,397], a step toward understanding the unwieldy number of possible microbial-molecule interactions across space and time. Given the wealth of data we collected on microbial and molecular communities, many interesting questions remain to be explored in this dataset. For example, by utilising co-occurrence networks[397,398], a powerful tool to understand ecological roles and interactions within biological communities, we can build a quantitative understanding of the connections between microbial taxa and the metabolites that exist along our environmental gradients[399].

5.2 The effect of forest harvest on the ecology of molecules in streams

There is ample evidence that harvest alters geochemical cycling in aquatic ecosystems following the removal of vegetation and disturbance of soils[135]. Yet, a general understanding of how these disturbances influence the ecology of molecules was not previously explored, and the mechanisms that could result in an altered ecology of molecules had not been described.

We explored how harvest impacted the ecology of molecules in streams. We found that by disrupting soil, harvest breaches spatial barriers that would otherwise isolate particular organic matter compounds from streams. Furthermore, by modulating catchment hydrology and introducing tree plant residue to the surface of soils, compounds otherwise rare in streams are introduced to the stream compound community as a short-lived but sizable pulse. These stream compound community changes persisted much longer than the increase in DOM concentrations. While the impacts of invasive species have been examined for decades, the introduction of novel molecules and the influence of these molecular-level changes on the abundance and diversity of compounds in the recipient stream ecosystems has not been explored. It follows that resulting temporal assembly and turnover in stream compound communities can play out in unexpected ways. Though, it is not yet clear how expectations should be formed. This influence of novel molecules would be an interesting avenue for future research. Moving forward, there is the potential to further explore the hydrological changes in these catchments by examining the chemodiversity produced by plant taxa in the catchment and utilising isotopic and hydrological tracers to develop mixing models based on molecular fingerprints. Furthermore, we collected FT-ICR-MS samples monthly for five consecutive months, from early summer to late fall. This data offers the opportunity to explore seasonal effects that inform the optimal timing of particular management efforts aimed at reducing the impacts of harvest on streams.

In **Chapter 4**, we expanded our investigation to a landscape level by presenting a PARAFAC model derived from a dataset of over 4500 samples with associated molecular characterization. We demonstrated that PARAFAC models could detect land-use disturbances, tree-species shifts, and wetland coverage in temperate and boreal forests. In particular, the PARAFAC component associated

with deeper soils and higher nitrogen concentrations was best predicted by harvest. This is consistent with the results from **Chapter 3** and prior studies of land-use changes performed in tropical forests[401]. This suggests that our observations may scale to the landscape level. Further work will be required to understand how biologically available this C is for stream microbial communities and if the magnitude or timing of delivery influences C dynamics.

We found that the most highly influential land cover variable predicting the composition of DOM was the relative proportion of particular tree species. This observation opens several lines of questioning. The first concerns phytochemical diversity and its variation across ecosystems[402]. Plants produce a wide range of metabolites which vary based on the plant species and environment. The result from this investigation suggests the molecular characteristics of DOM found in streams may reflect this plant metabolome. This could be an interesting investigation for high-resolution techniques to resolve. However, the bulk of the compound community may converge to a universal pool (**Chapter 2**), which suggests that only rare compounds may be useful in determining the impact of a particular tree or understory species. This may be an investigation worth performing given the expected changes in species range shifts under various climate change scenarios[403].

We also moved beyond previous literature and models with molecular associations by offering the first resolution of *unique* associations with each component and the redundancy amongst components. This offered novel interpretations for PARAFAC, as this model was particularly good at unambiguously resolving black carbon, polyphenols, and carboxyl-rich alicyclic molecules. Based on our results in **Chapters 2 and 3**, this can be applied to differentiate between surface and subsurface flow paths. Furthermore, there is clear potential utility in detecting the influence of harvest and interesting implications for investigations of drinking water quality and drought effects. We suggest practitioners use models like the one we developed, where molecular associations are defined if generating PARAFAC data in northern temperate and boreal forests. This dataset could be further explored by considering the proportion that each compound and compound grouping contributes to the relative signal intensity, as well as incorporating the impact of other land-use disturbances (e.g. insect outbreaks, fire, windthrow) and the spatial properties of water flow (e.g. hydrological weighting of impact areas).

5.3 Future directions and limitations

This dissertation was framed in the context of nature-based solutions for climate change and the concomitant increase in harvest pressure on forests. An important next step will be to develop holistic C budgets. Though we estimated that 8.5% of the C removed as wood during forest harvest was subsequently lost to waters (Chapter 3), comprehensive or large-scale C budgets (including losses from soils directly to the atmosphere[400]) were not in the scope of this thesis. However, there are interesting directions for further research. For example, globally, how much C is lost to freshwaters as a result of harvest activities on land? Furthermore, how do pulse effects such as storms and snowmelt (DOM accumulated over winter) influence the C budgets and DOM molecular composition, and how does seasonality play a role in constraining these budgets?

In Chapters 2 and 3, we performed an in-depth investigation of DOM along hillslopes flowing into streams. We inferred bio-availability based on incubation experiments paired with kinetic modeling[274,401] and the evidence of "what remains," i.e., compounds that are presumed to be difficult to degrade because they accumulate (or occur at the end of the land-ocean continuum or are of considerable age). However, investigating C use efficiency of particular compounds or mixtures, which ultimately determines the fate of the C in aquatic ecosystems, is needed to advance this research. Understanding the mechanisms that control outgassing and how changes in compound communities alter these is necessary to understand how our findings can be translated into action. Environmental context is also paramount for organic matter turnover[186], which suggests that the relationship between DOM molecular communities and C use efficiency in ecosystems must be investigated across multiple ecosystems[402–404]. Furthermore, in light of expected changes to climate, including altered temperature and precipitation and effects like recovery from acidification (browning), there are also expected shifts in tree species composition, understory, and likely the molecular composition of DOM in streams. There are many remaining questions and implications regarding climate change impacts on DOM.

The molecular diversity of natural organic matter has only recently become an analytical target for geochemistry. Most of the insights gained thus far, and the tool

we relied upon, are from the resolution of molecular formulae via ultrahigh-resolution mass spectrometry. However, many structural isomers contained within molecular formulae[97] can alter the function of a molecule in a particular environment. Emerging high-field nuclear magnetic resonance[405], multidimensional chromatography[406], and enzymatic assay approaches[407] must be further developed to gain insight into DOM structural diversity and how this influences the interpretation of FT-ICR-MS results. New theoretical frameworks and statistical approaches also need to be developed to fully appreciate the ecological data available from FT-ICR-MS. Network statistics and notions from statistical physics and complex systems theory may be required to identify properties of DOM-microbial systems across different ecosystems[62]. For example, some recent models do not prescribe reactivities to individual molecules a priori[408]. Instead, all compounds are continuously produced and consumed, and a dynamic equillibrium[186] controls their concentrations. These new methods may provide further insight into the complex nature of DOM in the forest-water nexus.

At present, active management of the DOM pool entering streams from forests remains out of reach and understanding the complex interactions amongst DOM components is in its infancy. There is a true paucity of process understanding of the interactions between DOM, microbes, and the environment, which presents an exciting frontier for those ready to explore the ecology of molecules and building such an understanding is especially important given the increasing demands on forests and waters. This thesis contributes to the small but growing body of literature that aims to build such an understanding.

Appendix A

A.1. Supplementary material

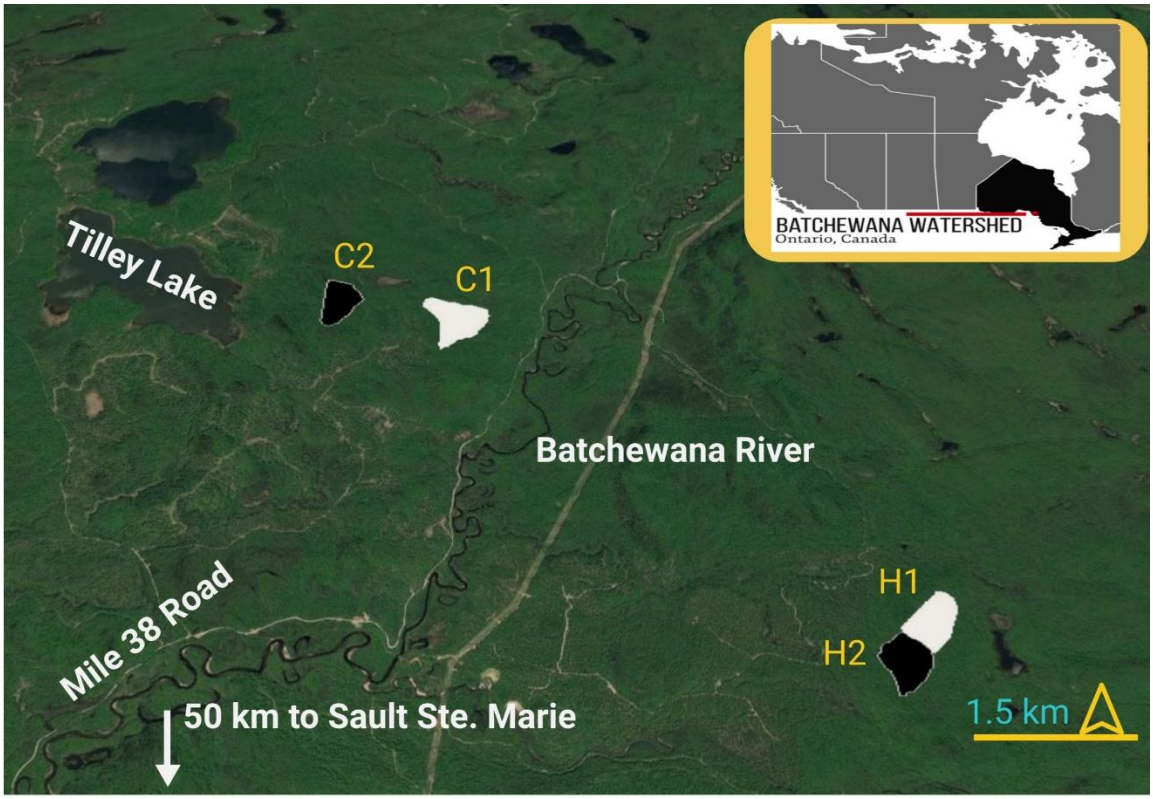


Fig. A.1: Study area within Batchewana watershed, central Ontario, Canada with four replicate catchments named C1, C2, H1, H2. These catchments were recruited as part of a forest harvest experiment and the names reflect this design: C = control and H = harvest. All samples were taken prior to these catchments being harvested and therefore can be treated as replicates. All geographic coordinates for sampling points can be accessed and downloaded from: https://www.gaiagps.com/public/VKMlEPlvf1t6ownd6Vulfm8T.

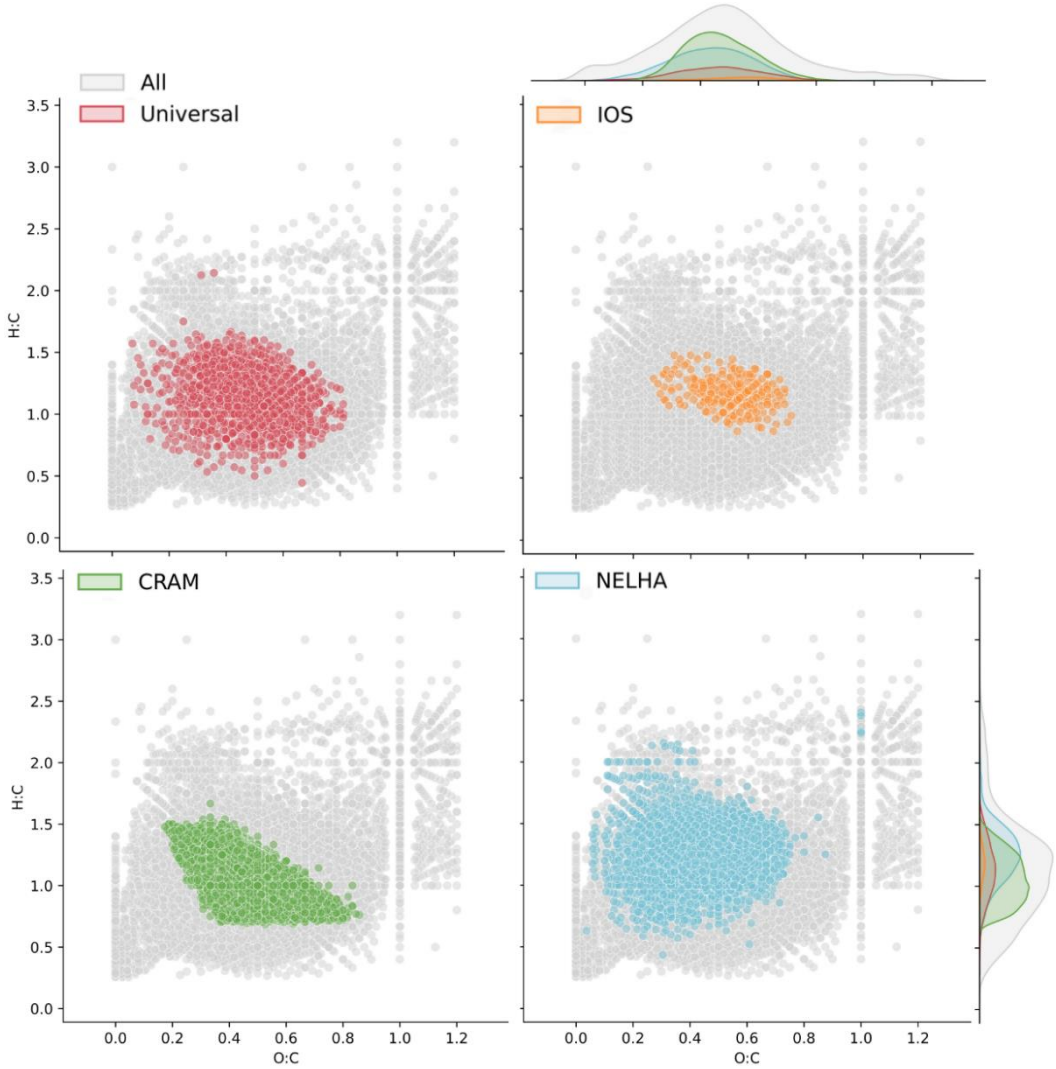


Fig. A.2: Molecular properties are shared by different definitions of degradation end-products that should be universally distributed. Each point is the elemental ratio (hydrogen to carbon and oxygen to carbon) of a molecular formula detected in our study ("All" category, n=9327). We further coloured these points according to whether they were present in every soil and stream sample in this study ("Universal"; n=1216), carboxyl-rich alicyclic molecules (CRAM, n = 3735), found in a subset of molecular formulae representing aged, degradation end-products identified in marine environments[158,409] (IOS, n=233), or found in a deep-sea reference sample from the Natural Energy Laboratory of Hawaii Authority (NEHLA) facility and expected to be dominated by degradation end-products (n=3019). CRAM were defined as molecular formulae that contained double bond equivalent (DBE) to C ratios of 0.30 to 0.68, DBE to H ratios of 0.20 to 0.95, and DBE to O ratios of 0.77 to 1.75[407]. Marginal axes show density curves for each compound category.

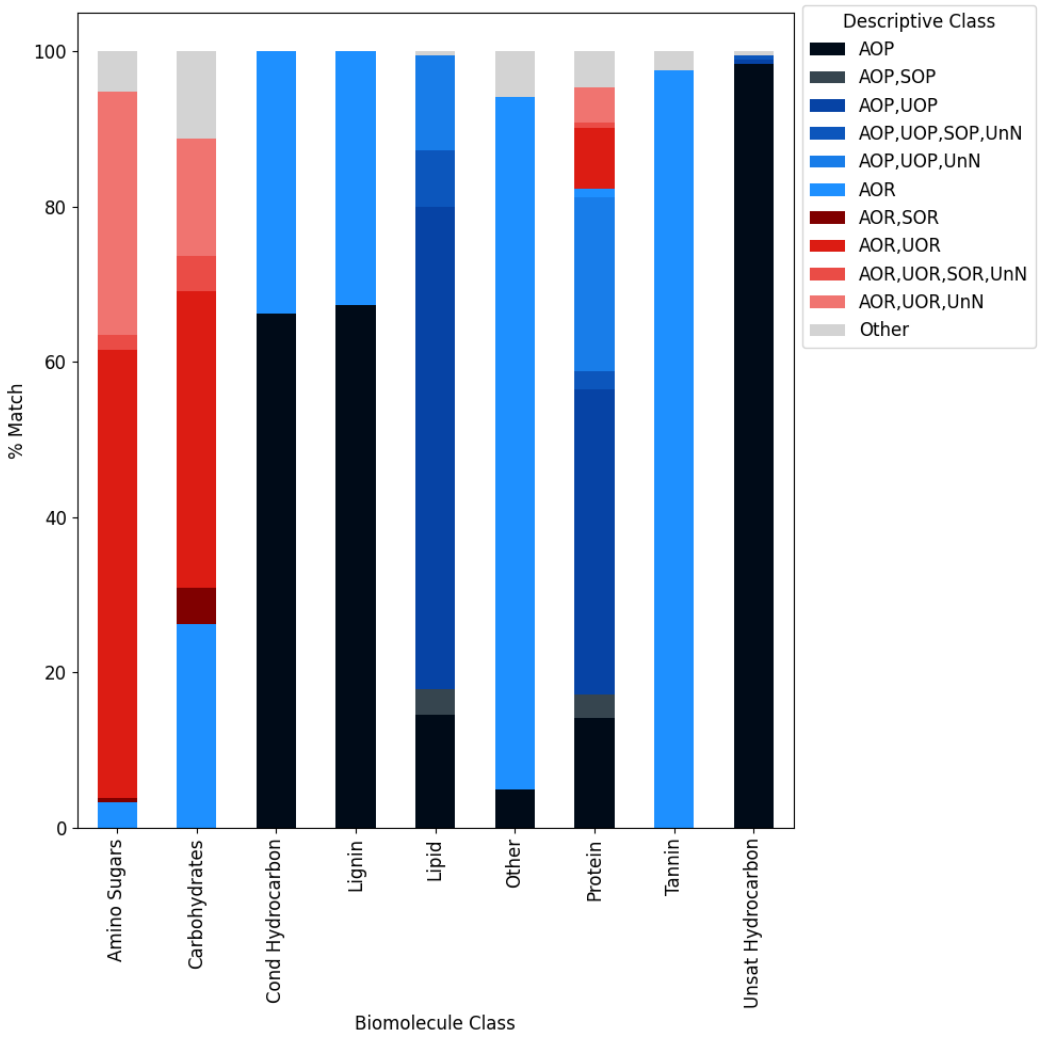


Fig. A.3: Comparison of compound classifications. We reclassified each molecular formula assigned to a compound class in the Main Text after ref. [4] using the descriptive classes of ref. [5]. Ref. [5] incorporates additional information on the modified aromaticity index and double bond equivalents of each formula in addition to the H:C and O:C ratios used by ref. [4]. The resulting descriptive classes can overlap and include: AOP = aromatic, oxygen-poor, SOP = saturated oxygen-poor, UOP = unsaturated oxygen-poor, UnN= unsaturated with N, SOR= saturated oxygen-poor or Other = unclassified.

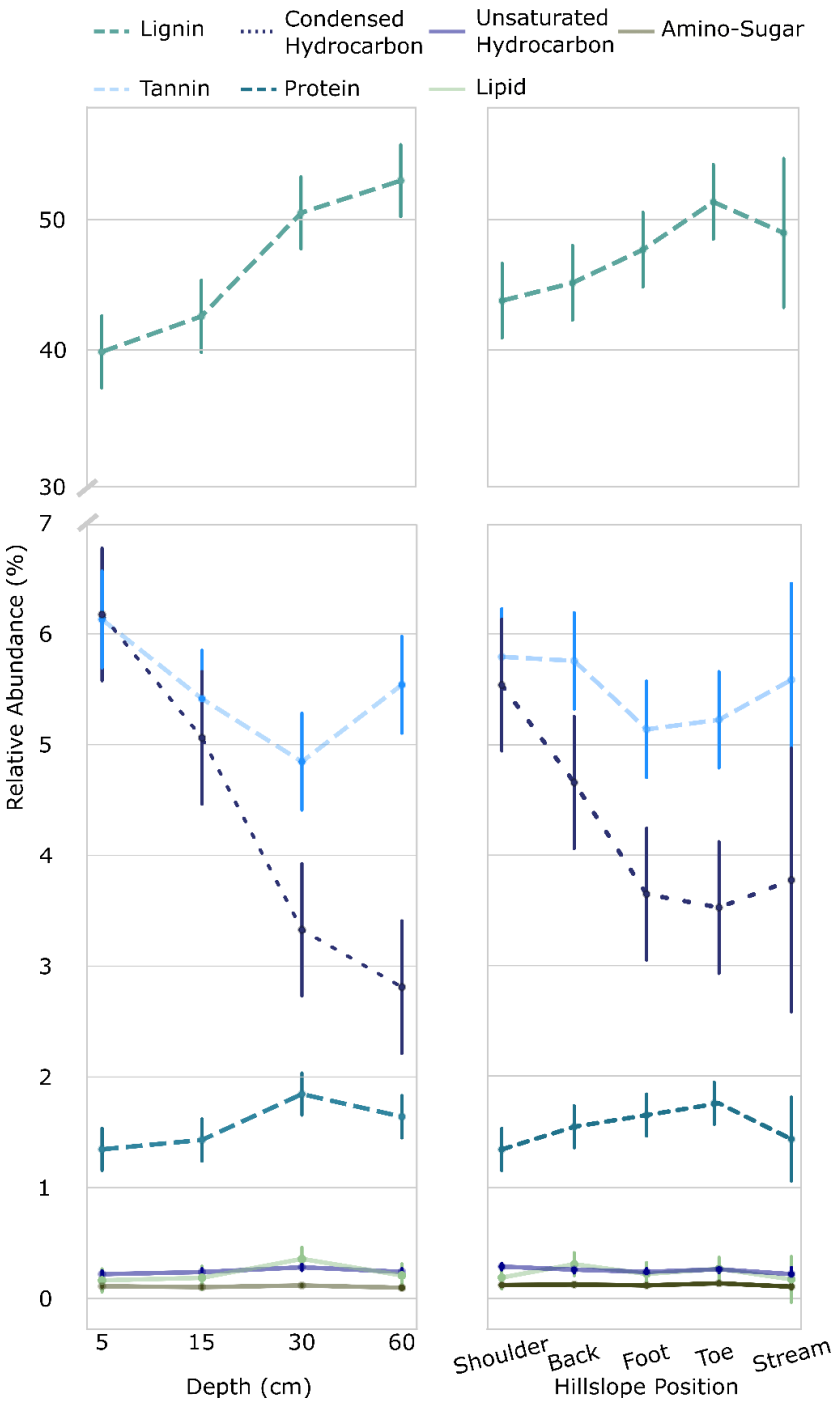


Fig. A.4: Compound classes within the universal pool shift in relative abundance across soil depth and hillslope position.Mean estimated molecular mass (± 95% CI) of DOM for each compound class when only universal compounds were considered. Molecular formulae were grouped into compound classes based on their atomic ratios. Darkened lines are those with statistically significant differences between 5 and 60 cm depths or shoulder and stream positions based on linear models (Table A4). The Tukey method was used to adjust p values for multiple comparisons.

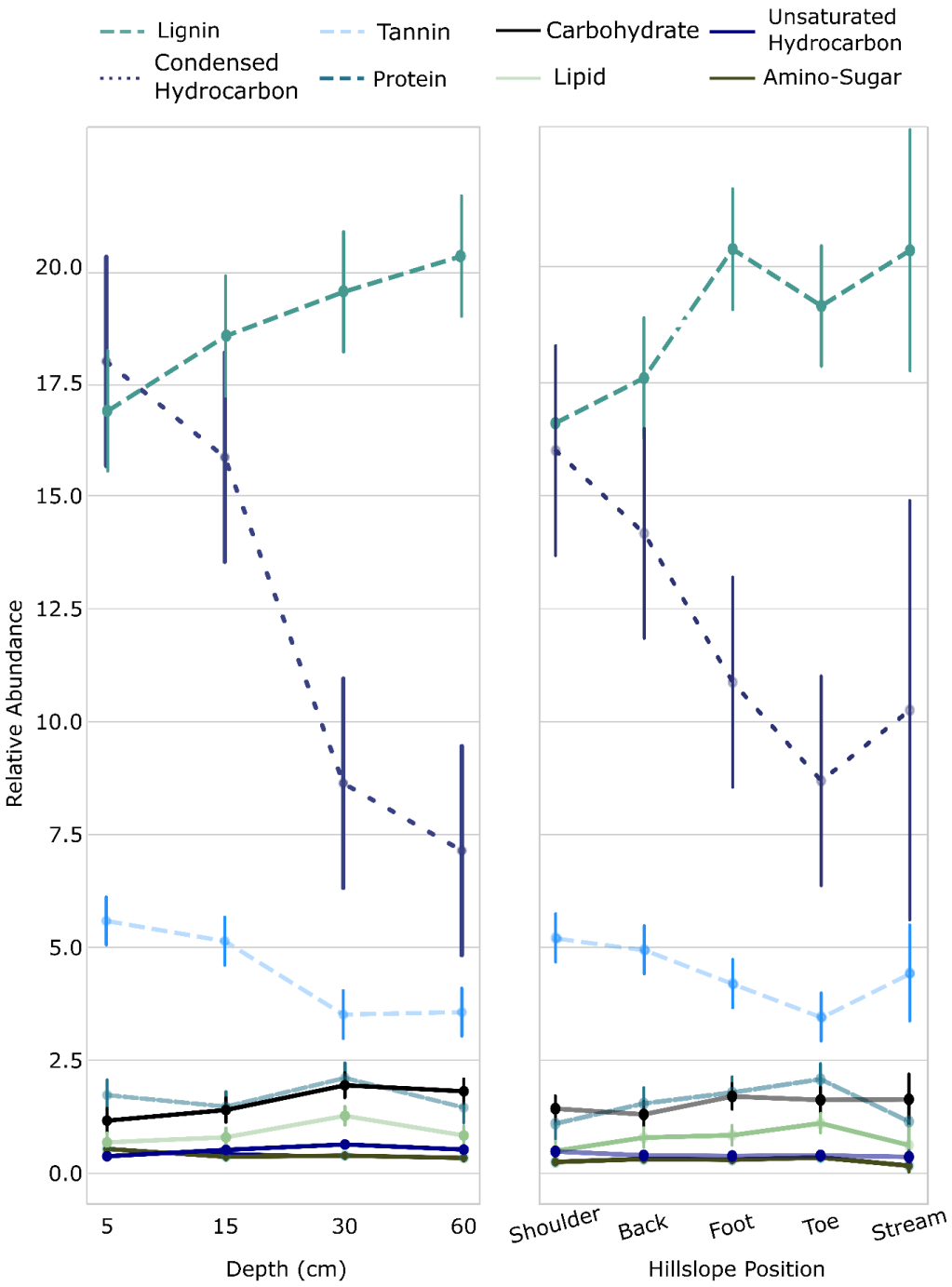


Fig. A.5: Compound classes within the non-universal pool shift in relative abundance across soil depth and hillslope position. Mean estimated molecular mass (± 95% CI) of DOM for each compound class when only universal compounds were considered. Molecular formulae were grouped into compound classes based on their atomic ratios. Darkened lines are those with statistically significant differences between 5 and 60 cm depths or shoulder and stream positions based on linear models (Table S5). The Tukey method was used to adjust p values for multiple comparisons.

Table A.1: Size-exclusion chromatography indicates high-molecular-weight substances (HMWS) outside the measurement size of ultra-high-resolution mass spectrometry were not abundant in our samples.CDOC = Dissolved organic carbon determined after chromatography. All measurements in mg C L^{-1}. "-" in HMWS concentrations indicates below limit of quantification (<0.5 mg C L^{-1}). % C pool refers to the % HMWS C in the CDOC pool.

Site	Hillslope Position	Depth (cm)	CDOC	HMWS	% C pool
C1	Shoulder	5	27.10	-	
C1	Shoulder	15	22.70	0.67	2.95
C1	Shoulder	30	17.70	-	
C1	Shoulder	60	26.57	1.49	5.61
C1	Backslope	5	40.20	0.52	1.29
C1	Backslope	15	35.90	-	
C1	Backslope	30	7.40	-	
C1	Backslope	60	10.60	-	
C1	Footslope	5	25.00	-	
C1	Footslope	15	9.90	-	
C1	Footslope	30	5.10	1.56	30.58
C1	Footslope	60	2.70	-	
C1	Toeslope	5	14.90	-	
C1	Toeslope	15	8.60	-	
C1	Toeslope	30	6.40	-	
C2	Shoulder	5	42.80	-	
C2	Shoulder	15	52.60	-	
C2	Shoulder	30	19.68	-	
C2	Shoulder	60	8.79	-	

C2	Backslope	5	31.80	-	
C2	Backslope	15	23.30	-	
C2	Backslope	30	17.60	1.2	6.82
C2	Backslope	60	6.70	-	
C2	Footslope	5	20.07	-	
C2	Footslope	15	11.30	-	
C2	Footslope	30	5.00	-	
C2	Footslope	60	3.60	-	
C2	Toeslope	5	7.80	-	
C2	Toeslope	15	3.80	-	
C2	Toeslope	30	5.90	-	
C2	Toeslope	60	1.50	-	
H1	Shoulder	5	15.10	-	
H1	Shoulder	15	20.87	-	
H1	Shoulder	30	9.35	-	
H1	Shoulder	60	8.30	-	
H1	Backslope	5	14.87	-	
H1	Backslope	15	14.40	-	
H1	Backslope	30	5.80	-	
H1	Backslope	60	4.70	-	
H1	Footslope	5	11.70	0.53	4.53
H1	Footslope	15	9.40	-	
H1	Footslope	30	4.30	-	
H1	Footslope	60	2.18	-	

H1	Toeslope	5	2.40	-	
H1	Toeslope	15	1.63	-	
H1	Toeslope	30	1.90	-	
H1	Toeslope	60	1.45	-	
H2	Shoulder	5	23.30	-	
H2	Shoulder	15	27.50	-	
H2	Shoulder	30	2.00	-	
H2	Shoulder	60	5.32	-	
H2	Backslope	5	14.20	-	
H2	Backslope	15	8.40	0.44	5.23
H2	Backslope	30	7.20	-	
H2	Backslope	60	8.30	0.53	6.39
H2	Footslope	5	3.54	-	
H2	Footslope	15	3.60	-	
H2	Footslope	30	3.45	-	
H2	Footslope	60	1.49	-	
H2	Toeslope	5	6.49	-	
H2	Toeslope	15	3.65	-	
H2	Toeslope	30	1.92	-	
H2	Toeslope	60	1.37	-	
HMWS detected (% samples)				13 (8 of 63)	

Table A.2: Linear models predicting the number and relative abundance of universal compounds. We modelled the proportion of molecular formulae that were counted in every

sample (n=1216), along with their relative abundance (i.e. summed normalised signal intensity). We compared these with the proportion of molecular formulae and relative abundance that overlapped with the "Island of Stability" (IOS, n=233) – a literature-based definition of universal-like molecular formulae[158,177,409] (Fig. S2). We also modelled the relative abundance of compounds in our samples that could be defined as carboxyl-rich alicyclic molecules (CRAM), which contain double bond equivalent (DBE) to C ratios of 0.30 to 0.68, DBE to H ratios of 0.20 to 0.95, and DBE to O ratios of 0.77 to 1.75[408] (n=3735), and compounds found in a deep-sea reference (n=3019) collected near the Natural Energy Laboratory of Hawaii Authority (NEHLA) facility, which should be dominated by widely distributed, degradation end-products. Values in cells are mean estimated effects ± standard error relative to the intercept, with the intercept expressed relative to zero. Proportions were modelled using a binomial error structure, so effects are expressed on a logit-scale and correspond with odds ratios. Bolded values were statistically significant at ***p < 0.001, **p < 0.01, *p < 0.05. Adjusted R^2 for linear models with binomial error structures were estimated using the MuMIn package in R[410]. For all, degrees of freedom = 57.

Coefficient	**Proportion of All Compounds**	**Abundance of All Compounds**	**Proportion IOS**	**Abundance of IOS**	**Abundance of CRAM**	**Abundance of NELHA**
Intercept: Site C1, 5 cm depth, shoulder position	**0.26 *** (0.03)**	**0.59 ** (0.06)**	**0.04 *** (0.00)**	**0.20 *** (0.02)**	**0.58 *** (0.04)**	**0.62 ** (0.11)**
Site C2	**1.19 * (0.10)**	0.53 (0.04)	**1.14 * (0.07)**	0.51 (0.03)	0.51 (0.02)	0.53 (0.06)
Site H1	1.12 (0.09)	0.51 (0.03)	1.09 (0.07)	0.52 (0.03)	**0.52 * (0.02)**	0.54 (0.06)
Site H2	1.05 (0.09)	0.50 (0.03)	1.05 (0.07)	0.52 (0.03)	**0.52 * (0.02)**	0.54 (0.06)
15 cm	1.00 (0.08)	0.51 (0.04)	1.00 (0.06)	0.52 (0.03)	0.52 (0.02)	0.54 (0.06)
30 cm	1.14 (0.10)	**0.57 *** (0.04)**	1.11 (0.07)	**0.59 *** (0.04)**	**0.56 *** (0.03)**	**0.65 *** (0.07)**
60 cm	**1.19 * (0.10)**	**0.60 *** (0.04)**	**1.14 * (0.07)**	**0.63 *** (0.04)**	**0.59 *** (0.03)**	**0.67 *** (0.07)**
Backslope	1.01 (0.09)	0.51 (0.04)	1.01 (0.07)	0.52 (0.03)	0.52 (0.02)	0.53 (0.06)
Footslope	0.99 (0.08)	0.51 (0.04)	0.99 (0.06)	**0.55 ** (0.04)**	**0.55 *** (0.03)**	**0.59 ** (0.06)**
Toeslope	1.16 (0.10)	**0.55 ** (0.04)**	1.12 (0.07)	**0.57 *** (0.04)**	**0.57 *** (0.03)**	**0.63 *** (0.07)**

Stream	0.90 (0.13)	**0.43 * (0.05)**	0.92 (0.10)	**0.38 *** (0.04)**	**0.40 *** (0.03)**	**0.33 *** (0.06)**
R^2	0.20	0.42	0.21	0.59	0.64	0.54

Table A.3: Linear models predicting the relative abundance of classes within the universal compound pool. Carbohydrates were not found in universal compounds and are omitted from this table. Values in cells are mean estimated effects ± standard error relative to the intercept, with the intercept expressed relative to zero. Bolded values were statistically significant at ***p < 0.001, **p < 0.01, *p < 0.05. For all, degrees of freedom = 57.

	Lignins	Tannins	Condensed Hydrocarbons	Proteins	Amino Sugars	Lipids	Unsaturated Hydrocarbons
Coefficient	*Estimates*						
Intercept: Site C1, 5 cm depth, shoulder position	**47.12 *** (2.92)**	**5.62 *** (0.47)**	**4.35 *** (0.65)**	**1.37 *** (0.21)**	**0.09 *** (0.01)**	0.12 (0.11)	**0.20 *** (0.03)**
Site C2	2.93 (1.84)	-0.16 (0.30)	-0.22 (0.41)	0.17 (0.13)	0.00 (0.01)	0.03 (0.07)	0.04 (0.02)
Site H1	1.43 (1.84)	-0.03 (0.30)	**-0.96 * (0.41)**	0.08 (0.13)	-0.01 (0.01)	0.03 (0.07)	0.00 (0.02)
Site H2	0.96 (1.84)	0.07 (0.30)	**-1.12 ** (0.41)**	-0.02 (0.13)	-0.01 (0.01)	0.09 (0.07)	-0.01 (0.02)
15 cm	2.66 (1.90)	**-0.72 * (0.31)**	**-1.11 * (0.42)**	0.08 (0.13)	-0.01 (0.01)	0.02 (0.07)	0.02 (0.02)
30 cm	**10.41 *** (1.90)**	**-1.28 *** (0.31)**	**-2.85 *** (0.42)**	**0.50 *** (0.13)**	0.01 (0.01)	**0.19 * (0.07)**	**0.06 ** (0.02)**
60 cm	**12.83 *** (1.90)**	-0.59 (0.31)	**-3.36 *** (0.42)**	**0.29 * (0.13)**	-0.01 (0.01)	0.04 (0.07)	0.02 (0.02)
Backslope	1.29 (1.90)	-0.04 (0.31)	-0.88 * (0.42)	0.21 (0.13)	0.00 (0.01)	0.12 (0.07)	-0.03 (0.02)
Footslope	3.68 (1.90)	**-0.66 * (0.31)**	**-1.90 *** (0.42)**	**0.31 * (0.13)**	-0.00 (0.01)	0.03 (0.07)	**-0.05 * (0.02)**
Toeslope	**7.14 *** (1.90)**	-0.57 (0.31)	**-2.02 *** (0.42)**	**0.42 ** (0.13)**	**0.02 * (0.01)**	0.08 (0.07)	-0.02 (0.02)
Stream	**11.37 *** (3.22)**	-0.86 (0.52)	**-3.60 *** (0.72)**	0.31 (0.23)	-0.02 (0.01)	0.05 (0.12)	-0.04 (0.04)
R^2	0.52	0.19	0.63	0.23	0.18	0.06	0.15

Table A.4: Linear models predicting the relative abundance of classes within the non-universal compound pool. Values in cells are mean estimated effects ± standard error relative to the intercept, with the intercept expressed relative to zero. Bolded values were statistically significant at ***p < 0.001, **p < 0.01, *p < 0.05. For all, degrees of freedom = 57.

Coefficient	Lignins	Tannins	Condensed Hydrocarbons	Proteins	Amino Sugars	Carbohydrates	Lipids	Unsaturated Hydrocarbons
Intercept: Site C1, 5 cm depth, shoulder position	**19.35 *** (1.38)**	**4.71 *** (0.58)**	**12.26 *** (2.52)**	**1.07 ** (0.37)**	**0.33 *** (0.08)**	**1.69 *** (0.30)**	**0.74 ** (0.23)**	**0.56 *** (0.07)**
Site C2	-0.86 (0.87)	-0.55 (0.36)	-1.80 (1.59)	0.19 (0.23)	-0.03 (0.05)	0.15 (0.19)	0.06 (0.14)	0.04 (0.05)
Site H1	**1.91 * (0.87)**	-0.35 (0.36)	-2.72 (1.59)	0.42 (0.23)	0.05 (0.05)	0.01 (0.19)	0.20 (0.14)	-0.01 (0.05)
Site H2	**3.00 ** (0.87)**	-0.26 (0.36)	**-3.55 * (1.59)**	0.44 (0.23)	**0.10 * (0.05)**	0.37 (0.19)	0.05 (0.14)	-0.04 (0.05)
15 cm	1.58 (0.90)	-0.45 (0.38)	-2.13 (1.64)	-0.26 (0.24)	**-0.18 *** (0.05)**	0.24 (0.20)	0.11 (0.15)	**0.14 ** (0.05)**
30 cm	**2.52 ** (0.90)**	**-2.08 *** (0.38)**	**-9.39 *** (1.64)**	0.39 (0.24)	**-0.15 ** (0.05)**	**0.80 *** (0.20)**	**0.60 *** (0.15)**	**0.27 *** (0.05)**
60 cm	**3.27 *** (0.90)**	**-2.03 *** (0.38)**	**-10.89 *** (1.64)**	-0.28 (0.24)	**-0.20 *** (0.05)**	**0.67 ** (0.20)**	0.16 (0.15)	**0.15 ** (0.05)**
Backslope	0.95 (0.90)	-0.26 (0.38)	-1.84 (1.64)	0.44 (0.24)	0.07 (0.05)	-0.13 (0.20)	0.29 (0.15)	-0.09 (0.05)
Footslope	**3.67 *** (0.90)**	**-1.01 ** (0.38)**	**-5.14 ** (1.64)**	**0.68 ** (0.24)**	0.05 (0.05)	0.27 (0.20)	**0.34 * (0.15)**	**-0.10 * (0.05)**
Toeslope	**2.47 ** (0.90)**	**-1.75 *** (0.38)**	**-7.33 *** (1.64)**	**0.97 *** (0.24)**	0.10 (0.05)	0.19 (0.20)	**0.61 *** (0.15)**	-0.09 (0.05)
Stream	**5.49 *** (1.52)**	**-1.92 ** (0.64)**	**-11.36 *** (2.78)**	0.01 (0.41)	**-0.22 * (0.08)**	0.63 (0.33)	0.35 (0.25)	0.02 (0.08)

R^2	0.43	0.50	0.56	0.27	0.29	0.24	0.30	0.32

Table A.5: Linear models predicting the intensity-weighted molecular mass of classes within the non-universal compound pool. Values in cells are mean estimated effects ± standard error relative to the intercept, with the intercept expressed relative to zero. Bolded values were statistically significant at ***p < 0.001, **p < 0.01, *p < 0.05. For all, degrees of freedom = 57.

Coefficient	Lignins	Tannins	Condensed Hydrocarbons	Proteins	Amino Sugars	Carbohydrates	Lipids	Unsaturated Hydrocarbons
Intercept: Site C1, 5 cm depth, shoulder position	**519.14 *** (12.17)**	**471.46 *** (15.51)**	**447.88 *** (18.92)**	**473.10 *** (16.79)**	**417.61 *** (19.66)**	**430.64 *** (18.47)**	**342.81 *** (12.66)**	**282.17 *** (10.87)**
Site C2	-6.26 (7.70)	-5.01 (9.81)	-8.30 (11.97)	-12.18 (10.62)	-7.81 (12.44)	-0.34 (11.68)	-4.23 (8.01)	-1.89 (6.88)
Site H1	**-18.34 * (7.70)**	7.16 (9.81)	-10.67 (11.97)	**-30.42 ** (10.62)**	-9.44 (12.44)	16.19 (11.68)	**-27.28 ** (8.01)**	4.53 (6.88)
Site H2	**-18.49 * (7.70)**	3.83 (9.81)	-2.27 (11.97)	-20.47 (10.62)	3.60 (12.44)	**28.44 * (11.68)**	**-18.83 * (8.01)**	10.69 (6.88)
15 cm	-2.29 (7.94)	7.15 (10.11)	-9.57 (12.34)	19.18 (10.94)	9.34 (12.82)	21.62 (12.04)	**27.93 ** (8.25)**	9.84 (7.09)
30 cm	**-17.56 * (7.94)**	-10.52 (10.11)	**-36.71 ** (12.34)**	-21.29 (10.94)	**-28.73 * (12.82)**	**54.31 *** (12.04)**	**34.96 *** (8.25)**	13.64 (7.09)
60 cm	-8.40 (7.94)	7.94 (10.11)	1.11 (12.34)	6.78 (10.94)	0.01 (12.82)	**72.43 *** (12.04)**	**31.79 *** (8.25)**	**36.91 *** (7.09)**
Backslope	-1.24 (7.94)	5.09 (10.11)	11.30 (12.34)	-4.02 (10.94)	3.46 (12.82)	20.29 (12.04)	15.70 (8.25)	**12.45 (7.09)**
Footslope	-8.12 (7.94)	12.82 (10.11)	3.65 (12.34)	**-24.83 * (10.94)**	1.77 (12.82)	**53.12 *** (12.04)**	4.87 (8.25)	**26.47 *** (7.09)**
Toeslope	**-25.69 ** (7.94)**	-0.23 (10.11)	-8.91 (12.34)	**-61.28 *** (10.94)**	-18.63 (12.82)	**65.37 *** (12.04)**	4.96 (8.25)	**21.86 ** (7.09)**

Stream	-11.00 (13.46)	18.84 (17.14)	12.18 (20.92)	9.43 (18.55)	16.31 (21.73)	**102.10 *** (20.42)**	**47.08 ** (13.99)**	**48.16 *** (12.02)**
R^2	0.21	-0.02	0.10	0.46	0.10	0.55	0.33	0.39

Table A.6: Linear models predicting the molecular mass of DOM. Molecular mass (Da) is equivalent to the mass to charge ratio (m/z) because all molecules are singly charged. Values in cells are mean estimated effects ± standard error relative to the intercept, with the intercept expressed relative to zero. Bolded values were statistically significant at ***p < 0.001, **p < 0.01, *p < 0.05. For all, degrees of freedom = 57.

Coefficient	**Intensity-Weighted Mass (All Compounds)**	**Mass (All Compounds)**	**Mass (Non-Universal Compounds)**
Intercept: Site C1, 5 cm depth, shoulder position	**437.94 *** (8.15)**	**476.77 *** (8.33)**	**491.24 *** (9.62)**
Site C2	-4.27 (5.16)	-7.07 (5.27)	-7.02 (6.08)
Site H1	-0.37 (5.16)	-7.19 (5.27)	-7.88 (6.08)
Site H2	1.47 (5.16)	-4.29 (5.27)	-4.32 (6.08)
15 cm	4.04 (5.31)	-1.03 (5.43)	-1.62 (6.27)
30 cm	-1.60 (5.31)	**-11.63 * (5.43)**	**-13.68 * (6.27)**
60 cm	10.70 * (5.31)	-3.53 (5.43)	-2.13 (6.27)
Backslope	4.77 (5.31)	2.90 (5.43)	3.93 (6.27)
Footslope	10.24 (5.31)	3.00 (5.43)	3.54 (6.27)
Toeslope	-0.29 (5.31)	-8.45 (5.43)	-9.85 (6.27)
Stream	**18.43 * (9.01)**	2.97 (9.21)	4.90 (10.63)
R^2	0.07	0.07	0.08

Table A.7: Linear models predicting changes in environmental variables at each depth and hillslope position at each site. Values in cells are mean estimated effects on a log-scale ± standard error relative to the intercept, with the intercept expressed relative to zero. Bolded values were statistically significant at ***p < 0.001, **p < 0.01, *p < 0.05. For all, degrees of freedom = 57.

Coefficient	DOC (mg L^{-1})	Bacterial Productivity (10^{-3} mg C L^{-1})	Humic Substances: Low Molecular Weight Substances	Carbon: Nitrogen (Total)
Intercept: Site C1, 5 cm depth, shoulder position	**4.98 *** (1.23)**	**0.01 *** (0.01)**	**2.33 * (0.80)**	**22.20 *** (7.16)**
Site C2	1.00 (0.13)	NA	1.12 (0.21)	0.72 (0.12)
Site H1	**0.51 *** (0.07)**	**2.40 * (0.95)**	**0.60 ** (0.11)**	**0.33 *** (0.06)**
Site H2	**0.52 *** (0.07)**	NA	0.74 (0.14)	**0.23 *** (0.04)**
15 cm	0.79 (0.11)	0.55 (0.30)	**1.50 * (0.28)**	**0.66 * (0.11)**
30 cm	**0.48 *** (0.06)**	0.38 (0.20)	0.91 (0.17)	**0.56 ** (0.10)**
60 cm	**0.28 *** (0.04)**	**0.12 ** (0.07)**	0.97 (0.19)	**0.46 *** (0.08)**
Backslope	**0.74 * (0.10)**	2.70 (1.41)	0.86 (0.16)	0.83 (0.14)
Footslope	**0.33 *** (0.04)**	**4.83 ** (2.63)**	**0.51 *** (0.09)**	**0.66 * (0.12)**
Toeslope	**0.20 *** (0.03)**	**0.02 *** (0.01)**	**0.39 *** (0.08)**	0.80 (0.14)
Stream	**8.40 *** (2.15)**	NA	**2.20 * (0.78)**	1.75 (0.58)
R^2	0.84	0.78	0.42	0.63

Table A.8: List of measured environmental variables. Variables were grouped by four categories (top row) corresponding to Fig. 4. Numbers in brackets following the CAZyme family name refer to, first, the number of individual CAZYme subfamilies included in our analyses and, second, the number that were important for explaining DOM composition across the 8 compound classes based on the redundancy analysis (RDA) inferential framework (see Methods). Bolded variables indicate the 62 environmental variables that were important for explaining DOM composition across the 8 compound classes. Bacterial and fungal taxonomic diversities were assessed using exact sequence variants (ESVs) generated by amplicon sequencing of the 16S rRNA gene and ITS2 region, respectively.

Expression of CAZymes (n=412, bold=51)	Enzyme Activity and Biomass (n=38, bold=6)	Microbial Diversity (n=6, bold=0)	Physical Chemistry (n=19, bold=4)
auxiliary activity family (14,1)	**2-hydroxy benzoic acid**	shannon diversity index (ITS)	**soil moisture content**
carbohydrate esterase (16,6)	4-hydroxy benzoic acid	Number of ESVs (16S)	**pH**
carboydrate-binding module (37,1)	**D, L-α-glycerol phosphate**	Number of ESVs (ITS)	conductivity
cohesin (1,0)	D-cellobiose	Shannon diversity index (16S)	alkalinity
dockerin (1,0)	D-dalactonic acid γ-lactone	Jaccard index (16S)	calcium
glycoside hydrolase (229,27)	D-galacturonic acid	Jaccard index (ITS)	potassium
glycosyl transferase (70,9)	D-glucosaminic acid		magnesium
polysaccharide lyase family (43,7)	D-malic acid		sodium
s-layer homology (1,0)	D-mannitol		sulphate

D-xylose		chlorine
glucose-1-phosphate		silicon dioxide
glycogen		nitrate
glycyl-L-glutamic acid		ammonium
itaconic acid		total inorganic carbon
L-Arginine		**aluminum**
L-Asparagine		iron
l-Erythritol		manganese
L-Phenylalanine		zinc
L-Serine		**total nitrogen**
L-Threonine		
N-acetyl-D-glucosamine		
N-acetylglucosaminidase		
phenylethylamine		
phosphatase		
putrescine		
pyruvic acid methyl ester		
tween 40		
tween 80		
xylanase		
α-cyclodextrin		

α-D-lactose		
α-ketobutyric acid		
β-glucosidase		
β-methyl-D-glucoside		
γ-hydroxybutyric acid		
microbial biomass		
basal respiration		

Table A.9: Differentially expressed CAZymes along the depth gradient. For each CAZyme, we calculated the mean number of normalised reads estimated over all the samples, along with the mean ± standard error (SE) of the effect of soil depth estimated from negative binomial generalized linear models using the R package DESeq2[411]. P-values from a Wald test were corrected for multiple comparisons with a Benjamini-Hochberg adjustment.

Carbohydrate-active enzyme	Mean count	Mean effect size	SE effect size	Adjusted p-value
Carbohydrate-Binding Module Family 8	96.21	-2.60	0.54	3.35 E-04
Polysaccharide Lyase Family 11	284.54	-0.53	0.11	3.78 E-04
Polysaccharide Lyase Family 1, Subfamily 2	910.32	-1.63	0.39	2.94 E-03
Glycoside Hydrolase Family 13 Subfamily 21	217.85	-0.95	0.24	5.53 E-03
Glycoside Hydrolase Family 5, Subfamily 54	91.11	-1.99	0.51	5.53 E-03
Glycoside Hydrolase Family 116	1476.49	-0.99	0.26	8.03 E-03
Glycoside Hydrolase Family 64	155.07	1.41	0.41	2.21 E-03
Glycoside Hydrolase Family 65	3457.61	-0.76	0.22	2.38 E-02
Carbohydrate Esterase Family 3	609.33	0.58	0.18	3.31 E-02
Glycoside Hydrolase Family 6	232.70	1.20	0.37	3.31 E-02
Carbohydrate-Binding Module Family 12	242.67	-1.08	0.34	3.78 E-02
Glycoside Hydrolase Family 20	4213.18	-0.57	0.18	3.78 E-02
Glycoside Hydrolase Family 17	950.16	0.68	0.22	3.40 E-02
Glycoside Hydrolase Family 73	192.13	0.54	0.17	3.40 E-02
Carbohydrate-Binding Module Family 48	1538.12	-1.28	0.42	4.18 E-02
Carbohydrate-Binding Module Family 6	750.91	1.53	0.51	4.58 E-02
Auxiliary Activity Family 2	3419.56	1.21	0.41	4.78E-02

Table A.10: Differentially expressed CAZymes along the hillslope gradient. For each CAZyme, we calculated the mean number of normalised reads estimated over all the samples, along with the mean ± standard error (SE) of the effect of soil depth estimated from negative binomial generalized linear models using the R package DESeq2[8]. P-values from a Wald test were corrected for multiple comparisons with a Benjamini-Hochberg adjustment. Genes explaining at least 2.5% of the variation in compound class are bolded.

Carbohydrate-active enzyme	Mean count	Mean effect size	SE effect size	Adjusted p-value
Auxiliary Activity 001	**32916.22**	**0.93**	**0.31**	**4.24E-24**
Polysaccharide Lyase 001, Subfamily 2	**910.32**	**1.28**	**0.39**	**1.18E-11**
Glycoside Hydrolase Family 13, Subfamily 26	**4268.76**	**-0.49**	**0.16**	**2.97E-10**
Glycoside Hydrolase Family 13, Subfamily 18	**335.14**	**1.02**	**0.34**	**3.90E-07**
Glycoside Hydrolase Family 51	**5063.26**	**1.07**	**0.30**	**1.44E-06**
Glycoside Hydrolase Family 135	**348.96**	**1.34**	**0.42**	**1.53E-05**
Glycoside Hydrolase Family 153	220.19	2.31	0.22	5.84E-05
Glycoside Hydrolase Family 127	2806.18	1.51	0.20	1.18E-04
Glycoside Hydrolase Family 44	5530.20	1.96	0.28	2.33E-04
Glycoside Hydrolase Family 125	398.41	1.90	0.32	2.57E-04
Glycosyltransferase Family 30	5307.16	1.13	0.20	2.57E-04
Glycosyltransferase Family 111	14.38	4.55	0.88	2.69E-04
Glycosyltransferase Family 21	49598.10	3.05	0.62	3.23E-04
Glycoside Hydrolase Family 65	3457.61	-1.05	0.22	3.58E-04
Polysaccharide Lyase Family 9, Subfamily 1	97.78	-1.31	0.29	3.58E-04
Glycoside Hydrolase Family 13, Subfamily 16	4526.45	-0.44	0.10	3.58E-04
Glycoside Hydrolase Family 5, Subfamily 7	329.90	0.87	0.19	1.02E-03
Glycoside Hydrolase Family 29	9898.17	1.66	0.37	1.34E-03
Glycoside Hydrolase Family 146	144183.44	3.32	0.75	1.42E-03
Carbohydrate-Binding Module Family 9	6125.91	1.43	0.33	1.42E-03
Glycosyltransferase Family 17	71.84	1.89	0.44	1.44E-03
Glycosyltransferase Family 5	10438.03	0.91	0.21	2.13E-03
Glycosyltransferase Family 22	353.05	-0.72	0.17	4.77E-03
Glycoside Hydrolase Family 3	28555.42	0.82	0.20	5.36E-03
Glycoside Hydrolase Family 140	5303.31	2.23	0.56	5.36E-03
Polysaccharide Lyase Family 6	62.10	-2.27	0.57	6.83E-03
Carbohydrate-Binding Module Family 50	3115.27	-0.35	0.09	7.02E-03
Glycoside Hydrolase Family 12	120.38	-0.67	0.17	7.02E-03
Glycoside Hydrolase Family 5, Subfamily 54	91.11	-1.88	0.52	7.18E-03
Glycoside Hydrolase Family 30, Subfamily 7	35.34	-1.18	0.33	9.47E-03

Polysaccharide Lyase Family 14, Subfamily 3	107.70	0.90	0.26	9.59E-03
Glycosyltransferase Family 107	91.25	0.81	0.23	1.03E-02
Polysaccharide Lyase Family 25	64.03	-3.05	0.87	1.11E-02
Glycosyltransferase Family 84	4586.90	0.86	0.25	1.13E-02
Glycoside Hydrolase Family 78	4423.06	0.61	0.18	1.17E-02
Glycoside Hydrolase Family 5, Subfamily 10	76.25	-0.63	0.19	1.17E-02
Glycoside Hydrolase Family 63	5970.43	0.85	0.25	1.19E-02
Glycoside Hydrolase Family 106	1733.42	0.69	0.21	1.44E-02
Glycoside Hydrolase Family 141	857.12	1.10	0.33	1.70E-02
Auxiliary Activity Family 3	42840.58	1.27	0.38	1.77E-02
Carbohydrate-Binding Module Family 48	1538.12	-1.37	0.42	1.95E-02
Glycoside Hydrolase Family 67	654.10	-0.32	0.10	1.95E-02
Glycoside Hydrolase Family 92	708.76	0.42	0.13	1.99E-02
Glycoside Hydrolase Family 37	2171.48	1.28	0.41	2.05E-02
Polysaccharide Lyase Family 11	284.54	-0.36	0.12	2.18E-02
Glycoside Hydrolase Family 144	3152.45	0.46	0.15	2.18E-02
Glycoside Hydrolase Family 16	130.03	-0.54	0.18	2.19E-02
Glycosyltransferase Family 4	120988.48	0.42	0.14	2.39E-02
Glycosyltransferase Family 7	40.61	1.25	0.42	2.66E-02
Glycosyltransferase Family 39	3342.74	-0.57	0.19	2.74E-02
Glycoside Hydrolase Family 87	306.14	0.91	0.31	3.05E-02
Carbohydrate-Binding Module Family 47	241.00	1.31	0.46	3.30E-02
Glycoside Hydrolase Family 43, Subfamily 12	137.27	-0.60	0.21	3.61E-02
Glycoside Hydrolase Family 13, Subfamily 13	401.51	-0.45	0.16	3.97E-02
Glycoside Hydrolase Family 47	1117.19	-1.15	0.41	3.97E-02
Glycoside Hydrolase Family 48	181.55	0.56	0.20	3.97E-02
Glycoside Hydrolase Family 5, Subfamily 17	54.05	-0.48	0.17	3.99E-02
Glycoside Hydrolase Family 43, Subfamily 28	103.16	-0.71	0.26	4.03E-02
Glycoside Hydrolase Family 5, Subfamily 1	154.20	-0.48	0.17	4.06E-02
Glycoside Hydrolase Family 133	5371.77	-0.48	0.18	4.13E-02
Carbohydrate-Binding Module Family 57	1376.46	0.52	0.19	4.35E-02

Table A.11: Data sources used to generate Figure 5. The mean percentage of molecules shared with a deep-sea reference sample was calculated for different sample types. If there was more than one study for a given sample type, the number of samples from each study is given in parenthesis.

Sample Type	Number of Samples	Study
5 cm Depth (Soil)	19	this one, Simon et al.[9] (3)
15 cm Depth (Soil)	20	this one, Simon et al.[9] (4)
30 cm Depth (Soil)	17	this one, Simon et al.[9] (1)
60 cm Depth (Soil)	17	this one, Simon et al.[9] (1)
Shoulder (Soil)	16	this one
Backslope (Soil)	16	this one
Footslope (Soil)	16	this one
Toeslope (Soil)	16	this one
Stream	15	this one (4), Hutchins et al. 2017[412] (11)
River	144	Hutchins et al. 2017[412]
Lake	116	Kellerman et al. 2014[274](115) Zark & Dittmar 2018[412] (1)
Bog	4	Simon et al. 2018[413]
Sea Surface	4	Zark & Dittmar 2018[414]
Aquifer	2	Simon et al. 2018[415]
Deep-sea	4	Simon et al. 2018[413]

Table A.12: Primer details for amplicon sequencing of the ITS2 and 16S rRNA regions.

JGI is the joint genome institute

Target gene	Primer name	Sequence	Reference
Fungal ITS2	ITS9F	GAACGCAGCRAAIIGYGA	Menkis et al., 2012[413,414] ; White et al., 1990[2]; used at JGI
	ITS4R	TCCTCCGCTTATTGATATGC	
Archaeal and Bacterial 16S rRNA V4-V5	515F-Y	GTGYCAGCMGCCGCGGTAA	Parada et al., 2015[415]; used at JGI
	926R	CCGYCAATTYMTTTRAGTTT	

(https://genome.jgi.doe.gov/portal/

Table A.13: Summary of physical and chemical environmental variables measured in soil pore water. D.L. = Detection limit.

Parameter/Range/ Detection Limit	Procedure	Principal Equipment
pH (3 - 9)	Orion Ross Ultra glass Electrode	Man-Tech PC-Titrate, Orion Thermo combination pH electrode
Specific Conductance (0 - 150 umho)	PCE-96-CT1003 electrode us/cm @ 25 C	Man-Tech PC-Titrate, with 4510 Conductivity meter
Total Alkalinity (0 - 2 meq/l)	Electrometric Titration	Man-Tech PC-Titrate, Titra-Sip titrator, Orion Thermo combination pH electrode
Total Nitrogen (0 - 2 ppm) D.L. - 0.05ppm	Automated Cadmium Reduction	Technicon Autoanalyser II - NO2+ NO3 channel - Autoclave Digestion - N.A.P. software
NH4 as N (0 - 500ppb) D.L. - 10ppb	Automated Sodium Nitroprusside, Filtered .45um	Seal Analytical AA3 Autoanalyzer AACE 6.07 software
NO2+NO3 as N (0 - 2ppm) D.L. - .04ppm	Automated Cadmium Reduction, Filtered .45um	Seal Analytical AA3 Autoanalyzer AACE 6.07 software
Total Phosphorus (unfiltered) (0 - 60ppb) D.L. - 1ppb	Automated Molybdophosphoric Blue	Technicon Autoanalyser II - N.A.P. software - Autoclave Digestion
Potassium – low end accuracy (water) 0.01ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Sodium - low end accuracy (water) 0.01ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Calcium – low end accuracy (water) 0.01ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Magnesium – low end accuracy (water) 0.01ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Sulphate (0 - 10ppm) D.L. -.2ppm	Conductance - ion exchange suppression	Dionex ICS 1100 Ion Chromatograph-Chromeleon 7.0 software
Total Sulphur as Sulphate	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument- Masshunter software

Chloride (0 - 5ppm) D.L. - .2 ppm	Conductance - Ion Exchange Suppression	Dionex ICS 1100 Ion Chromatograph-Chromeleon 7.0 software
Silica Oxide (0 - 50ppm) & (0-7ppm) D.L. - .25 ppm	Automated Ascorbic Acid	Technicon Autoanalyser II - N.A.P. software
Iron - low end accuracy (water) 0.005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Aluminum – low end accuracy (water) 0.005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Manganese - low end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Zinc – low-end accuracy (water) 0.001ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Copper - low-end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Nickel – low-end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Cadmium – low-end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Lead – low-end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Dissolved Organic Carbon (0 - 20 ppm) DL. - .4ppm	Sample Purged - N2 Acid & Potassium Persulphate - U.V. Rad. - Dialysis - O.C. Inversely Meas., Filtered .45um	Seal Analytical AA3 Autoanalyzer - AACE 6.07 software
Dissolved Inorganic Carbon (0 - 5ppm) DL. - .5ppm	Sample Acidified H2SO4 - CO2 Dialysis Thru Membrane - I.C. Inversely Meas., Filtered .45um	Seal Analytical AA3 Autoanalyzer - AACE 6.07 software
Soluble Reactive Phosphorus (0 - 60ppb) D.L. - 1ppb	Automated Molybdophosphoric Blue, Filtered .45um	Technicon Autoanalyser II - N.A.P. software
High-level NH4 as N (0 – 100ppm) D.L. – 2ppm	Automated Sodium Nitroprusside	Technicon Autoanalyser II - N.A.P. software
Arsenic – low-end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software

Cobalt – low-end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Chromium – low end accuracy (water) 0.0005ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Selenium – low end accuracy (water) 0.001ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software
Strontium – low-end accuracy (water) 0.001ppm	ICP-MS	Agilent 7700x Inductively Coupled Plasma Instrument - Masshunter software

Supplementary Results

To assess processing of DOM further, we averaged molecular masses in each sample weighted by the relative intensity of formulas[178]. In soils, increasing mass is thought to reflect the transformation of small plant-derived compounds into larger degradation products by microbes[178]. As expected if DOM shifted from plant- to microbial-derived compounds[178], the weighted molecular mass increased by an estimated mean (95% CI) of 2.5% (<0.1 to 5.1%) from 422 (415 to 430) Da to 433 (426 to 441) Da from 5 to 60 cm (Table S3). This shift was much stronger from shoulder into stream positions, increasing by 10.1% (4.2 to 16.0%) from 422 (414 to 430) Da to 437 (422 to 452) Da (Table S6). Increases in molecular mass were less consistent when compounds were not weighted by their relative intensities (Table S6). Furthermore, averaged values of all compounds lose the strong underlying signal seen when examining individual compound classes (see Main Text), though this averaged signal is comparable to the 21 Da (6% change) reported in Roth et al.[9].

Supplementary Methods

Rarefaction of DOM: We combined published data from Table S9 into a sample × molecular formula matrix. We then randomly generated 10,000 intervals between 250 to 20078, and randomly selected the corresponding molecular formula from the sample set. We then plotted the proportion of universal compounds in each sample as a function of the total number of sampled formulae, repeating this process 999 times. We visually identified a threshold of 6000 molecular formulae, after which additional sampling did not change the proportion of universal compounds per sample.

Bacterial productivity: In the laboratory, 1 g of soil from each sample was mixed with 50 mL distilled water using a gyratory shaker (200 rev min^{-1}) and soil particles were removed by centrifugation (1000 *g*) for 10 min. We then added 3.74 µL of ^{3}H-labeled leucine (1 mCi mL^{-1}, PerkinElmer, USA) to 1.5 mL of bacterial suspension. Blanks were prepared by immediately adding 75 µL of 100% trichloroacetic acid

(TCA), killing any live cells, and providing a measure of background ^{3}H leucine incorporation that was later subtracted from the incubated values. After 2 h of incubation at room temperature (~22°C), each sample received 75 µL of ice-cold 100% trichloroacetic acid (TCA) to stop incorporation. Samples were subsequently centrifuged at 13,000 *g* for 10 min and the supernatant was discarded. Each sample was then washed once with 5% TCA, 1.5 mL ice-cold 80% ethanol, and 0.2 mL 1M NaOH. Samples were vortexed and maintained at 90°C for 1 h. After cooling to room temperature, we added 1 mL of Ultima Gold scintillation fluid (Perkin Elmer, USA) to each sample and vortexed for 30 s. ^{3}H incorporation was measured with a LS6500 liquid scintillation counter (Beckman Coulter, USA).

Bacterial cell counts: Each sample was defrosted and we added 1.2 mL of a detergent: 250 mM tetrasodium pyrophosphate [TSP; pH 8.0] containing Tween 80, 0.5% final concentration. The solution was then vortexed for 30 s and shaken for 2 hr at 4°C. From the resulting slurry, 1 mL was layered slowly onto 0.5 mL of Histodenz solution (80% [wt/vol] prepared in 50 mM sterile TSP buffer). Bacterial cells and soil particles were separated by high-speed centrifugation (14,000 *g*) for 30 min. The upper and middle cell-containing phases (including the thin layer on top of the Histodenz cushion liquid phase) were carefully recovered and mixed with 1 mL of 50 mM TSP buffer and centrifuged at 17,000 *g* for 25 min. The supernatant was removed, and the cell pellet containing the cell fraction was resuspended in 0.8 ml of 50mM TSP buffer. The pellet slurries were vortexed and then stained with SYBR Green (final concentration 1×) and incubated for 20 min at room temperature in the dark. Samples were vortexed again before cytometry counts, and between each sample, 100 µL of sterile TSP buffer was passed through the flow cytometer.

Principal coordinates of neighbour matrices: The distance between the lysimeters was represented as a Euclidean distance matrix. Distances were calculated from the average latitude and longitude of the three replicate lysimeters at the same hillslope position. The resulting PCNM decomposes any spatial relationships between sample locations at the same site. This captures distinct information which justifies its inclusion alongside the ordinal variables (depth) and categorical variables (hillslope). We did not compute distances between lysimeters in different catchments by setting a threshold distance of 500 m[17].

Supplementary References

1. Kellerman, A. M. *et al.* Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. *Environ. Sci. Technol.* **52**, 2538–2548 (2018).
2. Behnke, M. I. *et al.* Pan-arctic riverine dissolved organic matter: Synchronous molecular stability, shifting sources and subsidies. *Global Biogeochem. Cycles* **35**, (2021).
3. Hertkorn, N. *et al.* Characterization of a major refractory component of marine

dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990–3010 (2006).

4. Kim, S., Kramer, R. W. & Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **75**, 5336–5344 (2003).
5. Merder, J. *et al.* ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Anal. Chem.* **92**, 6832–6838 (2020).
6. Lechtenfeld, O. J. *et al.* Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **126**, 321–337 (2014).
7. Barton. MuMIn: multi-model inference. R package version 1. 0. 0. *http://r-forge. r-project. org/projects/mumin/*.
8. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
9. Simon, C., Roth, V.-N., Dittmar, T. & Gleixner, G. Molecular Signals of Heterogeneous Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. *Front Earth Sci. Chin.* **6**, (2018).
10. Hutchins, R. H. S. *et al.* The optical, chemical, and molecular dissolved organic matter succession along a boreal soil-stream-river continuum. *J. Geophys. Res. Biogeosci.* **122**, 2892–2908 (2017).
11. Kellerman, A. M., Dittmar, T., Kothawala, D. N. & Tranvik, L. J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 3804 (2014).
12. Zark, M. & Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **9**, 3178 (2018).
13. Menkis, A. *et al.* Occurrence and impact of the root-rot biocontrol agent Phlebiopsis gigantea on soil fungal communities in Picea abies forests of northern Europe. *FEMS Microbiol. Ecol.* **81**, 438–445 (2012).
14. White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. in *PCR Protocols* 315–322 (Elsevier, 1990).
15. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
16. Roth, V.-N. *et al.* Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* **12**, 755–761 (2019).17. Borcard, D. & Legendre, P. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling vol. 153 51–68 Preprint at https://doi.org/10.1016/s0304-3800(01)00501-4 (2002).

Appendix B

B.1. Supplementary materials

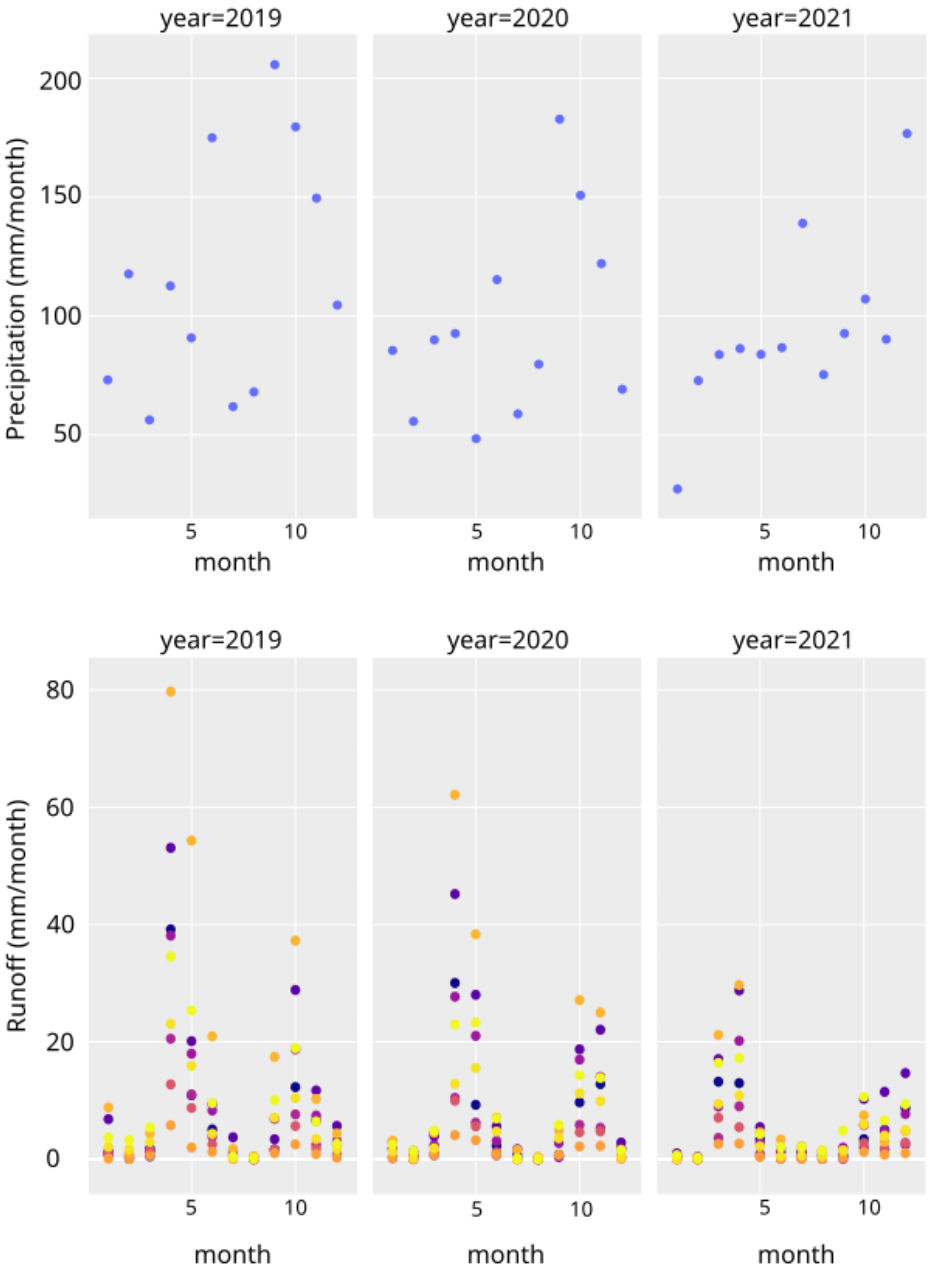


Fig. B.1 Precipitation and runoff in the Turkey Lakes Watershed for each year. *Runoff was measured in each of 9 catchments (individual colours).*

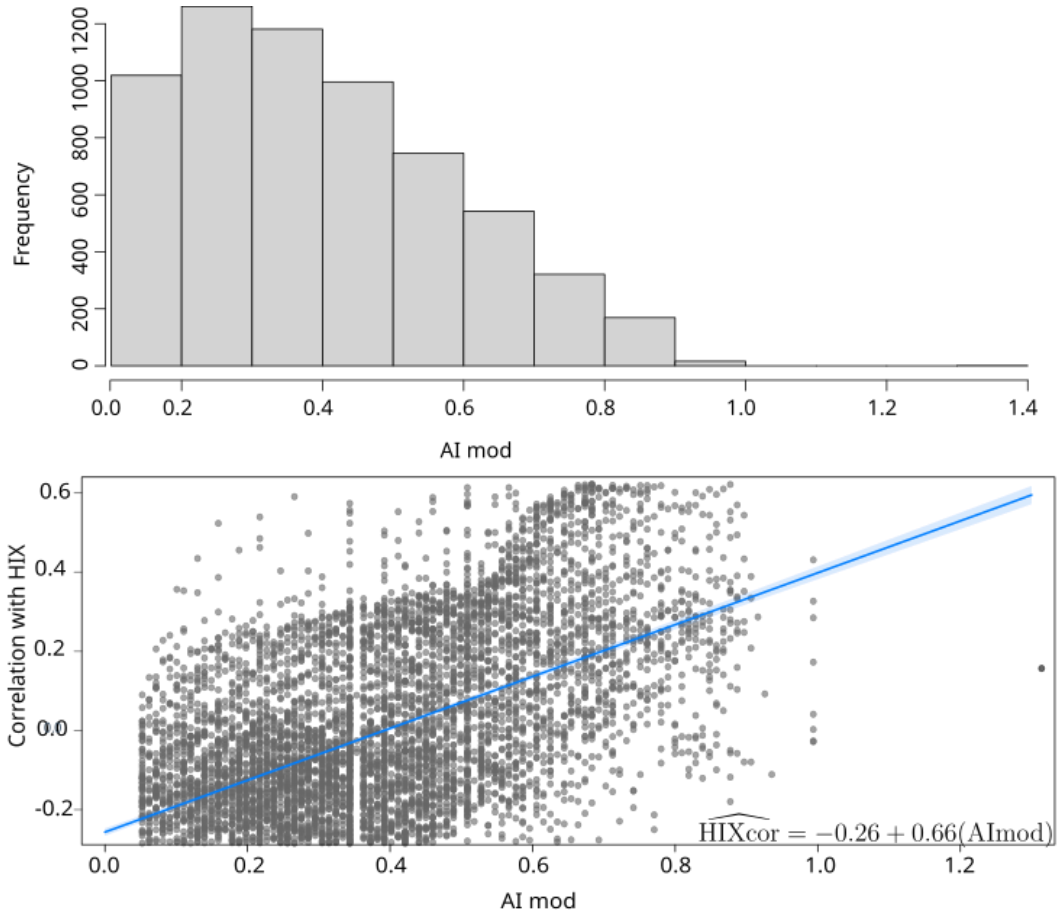


Fig. B.2 Most compounds are likely bioavailable based on their modified aromaticity (AI_{mod}). AI_{mod} in all molecular formulae (top) and their correlation with the humification index (HIX) (bottom). The reactivity of aromatic compounds (H:C< 1.1) is on average higher than reactivity in the "highly unsaturated" region (1.1 < H:C < 1.5)[418] and the uptake rate of DOM first decreases with increasing AI_{mod}, but at AI_{mod} between 0.25 and 0.33 began to increase again.

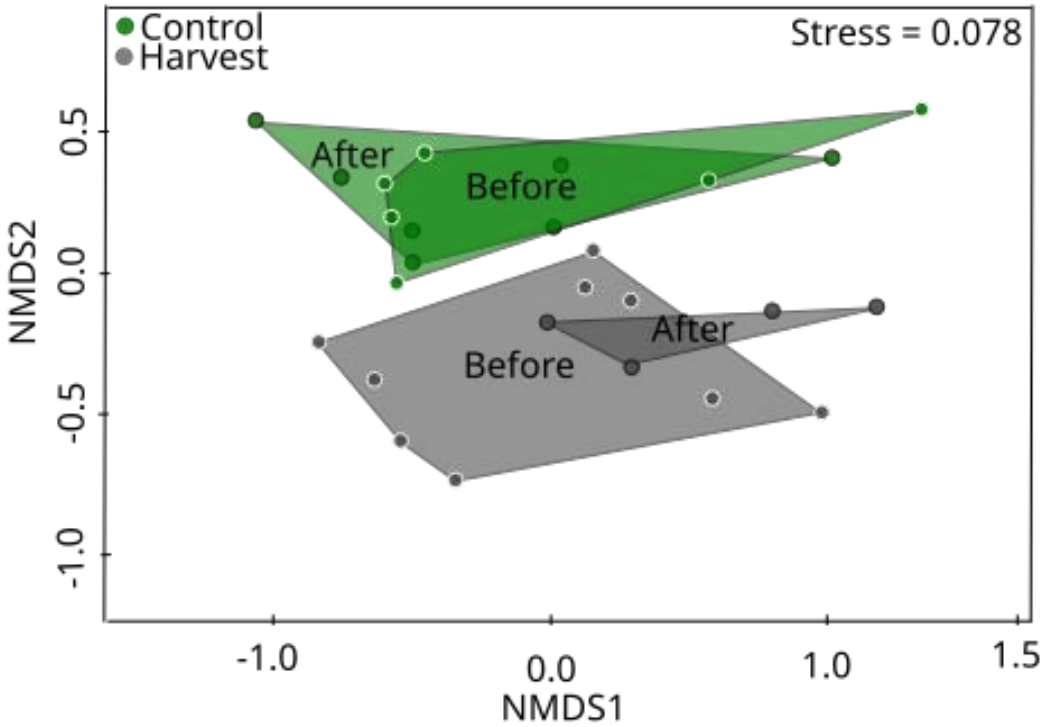


Fig. B.3 Non-metric multidimensional scaling (NMDS) ordination of DOM composition in streams before (June-September 2020, n=9) and after (October-November 2020, n=4) logging in replicate (n=4) catchments. The NMDS was fitted by calculating the Jaccard distance between observations from the presence-absence of molecular formulae. Compositional differences among groups were identified using a permutational multivariate analysis of variance (R^2 = 0.213 p = 0.029). Polygons were created by connecting the outermost data points within a grouping.

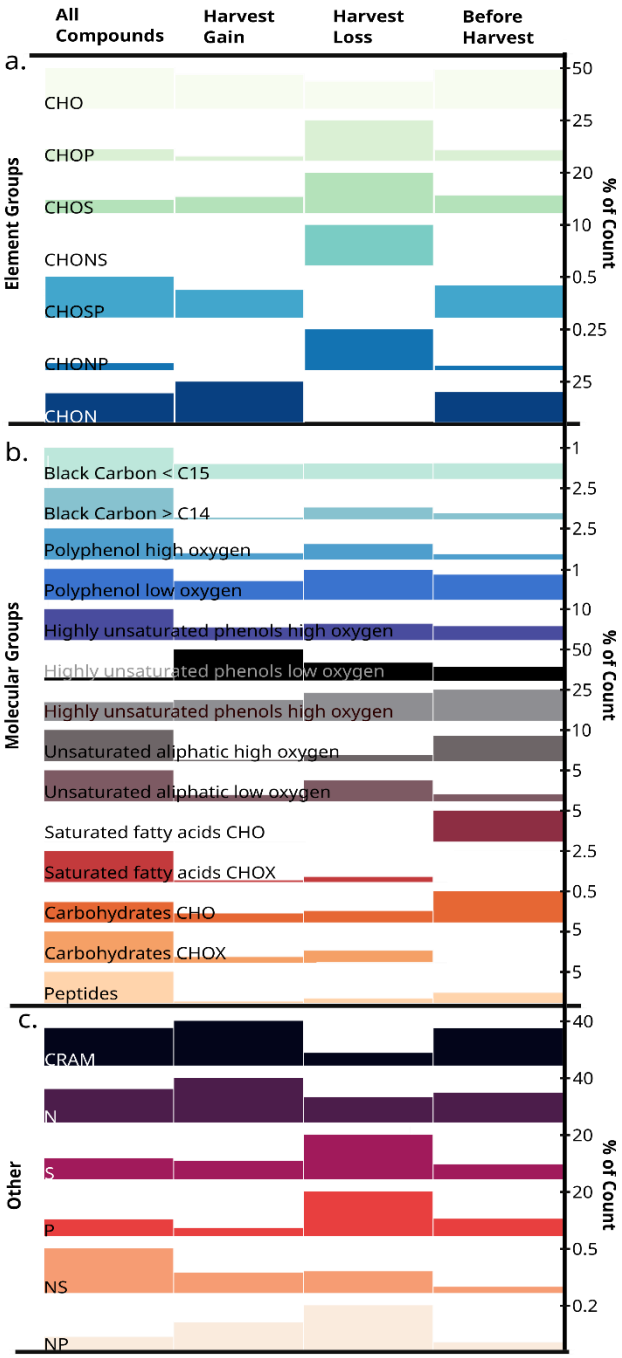


Fig. B.4 Composition of DOM classified into different categories. We classified all compounds (n=7444) based on **a.** the elements they contained, **b.** Groupings based on stoichiometry relating to compound structure **c.** Carboxyl-rich alicyclic molecules (CRAM) and elemental groups that sum non-exclusively. We separately classified compounds that appeared in the harvest site only after logging (n=1035); compounds that were lost from the harvest site after logging (n=320); and all compounds present in the harvest site before logging (n=4927). Ticks are aligned to the highest bar in each row. The base of the bars is

0%. Columns in **a**-**b** sum to 100%, while in **c**. the groupings are not mutually exclusive and can exceed 100%. Black Carbon <15 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C<15. Black Carbon >C14 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C >= 15. Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. Highly unsaturated phenols high oxygen= O-rich highly unsaturated compounds, soil-derived humics, phenolics. Highly unsaturated phenols low oxygen =O-poor highly unsaturated compounds, soil-derived humics, phenolics. Unsaturated aliphatics oxygen rich = O-rich unsaturated aliphatic compounds, Unsaturated aliphatic oxygen poor= O-poor unsaturated aliphatic compounds, Saturated fatty acids CHO = saturated fatty acids CHO, without N, S or P, Saturated fatty acids CHOX = saturated fatty acids CHOX, with N, S or P. Carbohydrates CHO = Carbohydrates, sugars without N, S, or P. Carbohydrates CHOX = carbohydrates, sugars with N, S, or P. Peptides= peptides (unsaturated aliphatic and with at least 1 N).

Table B.1 Estimated effects for models predicting the dissolved organic matter concentration in streams in the harvest year (2020). Values in cells are the mean estimated effect relative to the intercept, with the intercept expressed relative to zero. Bolded effects do not overlap zero. CI are the 95% confidence interval. Model predictors were time (in days), Period (Before or After), and Treatment (Control or Harvest). Site was included as a random effect. σ is the within-site variance and τ is the between-site-variance. R-squared values are provided as marginal (considering only fixed) and conditional R-squared statistics (considering both fixed and random effects)[2].

Fixed effects	Estimates	CI
Intercept (Before, Control)	**6.54**	5.32 – 7.77
Time (days)	**0.02**	0.01 – 0.03
After	-0.25	-1.44 – 0.95
Treatment	**-5.28**	-8.56 – -2.01
Time × Before/After	**-0.07**	-0.11 – -0.03
Time × Control/Harvest	**-0.02**	-0.04 – -0.00
Before/After × Control/Harvest	0.69	-0.90 – 2.29
Time × Before/After × Control/Harvest	**0.08**	0.01 – 0.14
Random Effect Estimates		
σ	0.41	
τ	0.08	
N Sites	4	
Observations	27	
Marginal R^2 / Conditional R^2	0.86 / 0.91	

Table B.2 Estimated effects for models predicting the chemistry and fluorescence variables in streams in the harvest year (2020).

Model predictors were time (in days), period (Before or After), and treatment (Harvest or Control). Values in cells are mean estimated effects (standard error) relative to the intercept in parentheses, with the intercept expressed relative to zero. Bolded values were statistically significant at ***p <0.001, **p < 0.01, *p < 0.05. Site was included as a random effect to control for catchment-specific differences, σ is the within-site variance and τ is the between-site-variance. R-squared values are conditional R-squared statistics (taking both fixed and random effects into account)[416]. We also tested the fluorescence index (FI)[417], biological index (BIX)[418], quotient of molar absorptivity at 280 nm (E2E3)[419], and spectral slopes at 275–295 nm (S275–295), 350–350 nm (S350–400) and the spectral slope ratio (SR), absorbance at 254 and 300 (a254, a300), and concentrations of: total organic carbon (TOC), dissolved inorganic carbon, potassium, aluminum, iron, zinc, cadmium, nickel, and copper. None of the estimated effects in these models were significant. All additional absorbance measures were measured with the same methods as HIX (see Methods).

	Fixed Effects						**Random Effects**		
Response	**Intercept**	**Time**	**BA**	**CI**	**BA × CI**	**Time × BA × CI**	**σ**	**τ**	**R^2**
NO_3	-0.01 (0.09)	**-4.57** (1.88)	0.02 (0.10)	**0.57** (0.11)	-0.15 (0.13)	-2.28 (4.89)	0.01	0.75	0.92/ 0.93
NH_4	-0.00 (0.01)	**-0.34** (0.13)	0.01 (0.01)	0.01 (0.01)	-0.01 (0.01)	-0.41 (0.34)	0.03	0.05	0.34/ 0.39
K	0.09 (0.14)	**-8.27** (2.65)	**0.41** (0.13)	0.37 (0.17)	**-0.47** (0.16)	-2.09 (6.60)	0.02	0.75	0.42/ 0.42
Cl	0.04 (0.04)	-1.52 (0.75)	**0.13** (0.04)	0.07 (0.04)	-0.10 (0.05)	-0.23 (1.95)	0.06	0.09	0.39/ 0.45
Al	0.07 (0.04)	-0.79 (0.72)	**0.10** (0.04)	-0.06 (0.04)	-0.08 (0.04)	0.30 (1.83)	0.09	0.12	0.78/ 0.79
Fe	0.01 (0.04)	-1.24 (0.65)	**0.08** (0.03)	-0.00 (0.05)	**-0.08** (0.03)	-0.63 (1.53)	0.07	0.22	0.37/ 0.40

	Fixed Effects						Random Effects		
Response	**Intercept**	**Time**	**BA**	**CI**	**BA × CI**	**Time × BA × CI**	**σ**	**τ**	**R^2**
Zn	11.03 (4.99)	126.03 (99.62)	-3.07 (4.89)	-7.80 (6.12)	1.26 (6.13)	**660.22** (251.46)	20.45	5.53	0.27/ 0.32
Cd	0.14 (0.09)	-1.55 (1.88)	-0.19 (0.10)	-0.00 (0.11)	0.11 (0.13)	**-14.10** (4.87)	0.01	0.09	0.62/ 0.65
Ni	0.04 (0.16)	**-7.85** (3.29)	0.08 (0.18)	0.13 (0.19)	-0.13 (0.23)	-8.66 (8.52)	0.02	0.75	0.44/ 0.49
HIX	**18.71** (2.15)	34.53 (25.78)	-5.75 (2.76)	**-13.54** (2.73)	**8.98** (3.83)	-106.73 (157.37)	7.14	0.11	0.62/ 0.79

Table B.3 Estimated effects for models predicting the chemistry and fluorescence variables in streams between 2019-2021. Model predictors were time (in days), Period (Before or After), and Treatment (Harvest or Control). Values in cells are mean estimated effects ± standard error relative to the intercept in parentheses, with the intercept expressed relative to zero. Bolded values were statistically significant at $p < 0.05$. Site and Year were selected as random effects to control for catchment-specific differences within year. σ is the within-site variance, τ1 is the between-group-variance by year, and τ2 is the between-group-variance by site. R-squared values are marginal R-squared statistics (taking only fixed effects into account). We also tested the fluorescence index (FI)[417], biological index (BIX)[418], quotient of molar absorptivity at 280 nm (E2E3)[419], and spectral slopes at 275–295 nm (S275–295), 350–350 nm (S350–400) and the spectral slope ratio (SR), absorbance at 254 and 300 (a254, a300), and concentrations of: total organic carbon (TOC), dissolved inorganic carbon, potassium, aluminum, iron, zinc, cadmium, nickel, and copper. None of the estimated effects in these models were significant. All additional absorbance measures were measured with the same methods as HIX (see Methods).

	Fixed Effects						**Random Effects**			
Response	**Intercept**	**time**	**BA**	**CI**	**BA × CI**	**Time × BA × CI**	**σ**	**τ 1**	**τ 2**	**R^2**
NO_3	**0.18** (0.09)	0.40 (0.37)	**-0.13** (0.06)	0.48 (0.12)	-0.15 (0.09)	1.19 (0.77)	0.01	0.00	0.10	0.93
pH	**6.88** (0.17)	0.07 (0.39)	**-0.25** (0.12)	0.02 (0.23)	0.02 (0.17)	**1.48** (0.67)	0.03	0.19	0.04	0.49
Conductivity	**37.08** (7.71)	**40.16** (11.33)	**-14.91** (3.46)	16.69 (10.82)	-3.43 (4.80)	28.69 (20.21)	23.84	10.12	0.87	0.85
Alkalinity	**0.25** (0.07)	**0.31** (0.14)	**-0.09** (0.04)	0.15 (0.10)	-0.05 (0.06)	0.13 (0.25)	0.92	0.09	0.78	0.71
Ca	**4.74** (0.96)	**4.10** (1.69)	**-1.90** (0.49)	1.79 (1.34)	-0.45 (0.69)	2.17 (3.12)	0.49	1.21	0.20	0.76

	Fixed Effects						Random Effects			
Response	**Intercept**	**time**	**BA**	**Cl**	**BA × Cl**	**Time × BA × Cl**	**σ**	**τ 1**	**τ 2**	**R^2**
Mg	**0.79** (0.26)	**0.72** (0.34)	**-0.26** (0.10)	0.52 (0.37)	-0.24 (0.14)	0.60 (0.62)	0.02	0.35	0.04	0.81
Na	**1.04** (0.26)	0.60 (0.34)	**-0.25** (0.11)	0.73 (0.37)	**-0.31** (0.15)	0.34 (0.59)	0.02	0.35	0.38	0.85
SO_4	**2.88** (0.45)	2.46 (1.23)	**-1.23** (0.35)	0.14 (0.62)	0.84 (0.48)	1.49 (2.31)	0.22	0.48	0.11	0.54
Mn	**11.40** (2.90)	**24.88** (11.76)	-5.76 (3.57)	-8.15 (3.88)	5.83 (4.88)	9.35 (21.04)	23.16	1.11	0.10	0.27
Pb	**0.13** (0.03)	0.00 (0.14)	**-0.10** (0.04)	-0.05 (0.04)	0.04 (0.06)	-0.20 (0.25)	0.87	0.17	0.80	0.38
HIX	**18.37** (2.22)	34.10 (24.17)	**-5.54** (2.27)	**-14.24** (2.72)	**8.57** (2.78)	28.16 (26.82)	6.27	0.00	1.36	0.78

Table B.4 Compound counts in treatments. All values were estimated by mixed effects models. CI is the 95% confidence interval.

Treatment	*Estimates*	*CI*
Before, Control	**5386.86**	4733.01 – 6131.57
Before, Harvest	**4771.52**	3999.31 – 5692.08
Control, After	**4152.74**	364095 – 4738.82
Harvest, After	**6292.24**	5521.74 – 7172.08
Model statistics		
N Site	4	
Observations	24	
Marginal R^2	0.99	

Table B.5 Pairwise dissimilarity between matched control sites and harvests before and after logging treatment. All values were estimated by a linear model with site as a fixed effect to control for site-level differences. Bolded effects do not overlap zero. CI is the 95% confidence interval.

Predictors	*Estimates*	*CI*
Before	**0.11**	0.02 – 0.19
After	**0.12**	0.00 – 0.24
R^2 / R^2 adjusted	0.21 / 0.16	

Table B.6 Percentages of each compound class of the total number of compounds in each group of compounds. The total number of molecules is given in parentheses. We separately classified compounds into groups that appeared in the harvest site only after logging (harvest gains) and all compounds present in the harvest site before logging.

Group (n)	CHO	CHOP	CHOS	CHONS	CHOSP	CHONP	CHON
Harvest Gains (1035)	45.12	3.03	8.40	0.20	0.49	0.00	42.77
Harvest Losses (320)	36.33	24.1	20.26	18.65	0.00	0.32	0.32
All (7444)	51.39	6.63	9.01	0.33	0.57	0.04	32.02
Before Harvest (4927)	54.00	7.36	6.99	0.06	0.71	0.06	30.82

Table B.7 Percentages of each elemental class of the total number of compounds in each group of compounds. The total number of molecules is given in parentheses. Elemental classes are not mutually exclusive; therefore, rows can sum to >100. We separately classified compounds that appeared in the harvest site only after logging (harvest gains); compounds that were absent from the harvest site after logging (harvest losses); all compounds present in all streams (all) and all compounds present in the harvest site before logging.

Group (n)	N	S	P	NS	NP
Harvest Gains (1035)	42.00	9.18	3.67	0.29	0.19
Harvest Losses (320)	24.38	21.88	19.06	0.31	0.31
All (7444)	32.28	10.45	7.27	0.63	0.09
Before Harvest (4927)	28.76	7.53	7.54	0.09	0.06

Table B.8 The difference in mean soil-stream similarity before and after harvest in treatment and control sites. Estimated mean (± 95% CI) pairwise similarity between soil and stream sites before (intercept) and after logging treatment in each of the control and harvest. CI are the 95% confidence interval.

Harvest	**Similarity (Soil-Stream)**	
Predictors	*Estimates*	*CI*
Before Harvest	**0.61**	0.55 – 0.67
After Harvest	-0.04	-0.11 – 0.03
Observations	41	
Control	***Similarity (Soil-Stream)***	
Before Harvest	**0.60**	0.58 – 0.62
After Harvest	**-0.07**	-0.10 – -0.04
Observations	42	

Table B.9 The difference in soil (depth x position)-stream similarity before and after harvest in the treatment and controls. Estimated mean (± 95% CI) pairwise similarity between soil and stream sites after logging treatment (compared to before) in each of the control and harvest, at each soil position (depth **×** position**)**. CI are the 95% confidence interval. Bolded effects do not overlap zero.

Predictors	*Estimates*	*CI*
Shoulder, 5 cm	-0.08	-0.79-0.11
Shoulder, 15cm	**0.20**	**2.23-0.09**
Shoulder, 30 cm	-0.02	-0.23-0.12
Shoulder, 60cm	-0.03	-0.22-0.13
Backslope, 5 cm	0.09	0.09-0.95
Backslope, 15cm	-0.09	-0.57-0.17
Backslope, 30cm	-0.02	-0.21-0.07
Backslope, 60cm	**0.13**	**0.57-2.31**
Footslope, 5cm	0.08	0.09-0.91
Footslope, 15cm	**0.25**	**0.09-2.74**
Footslope, 30cm	0.03	0.13-0.27
Footslope, 60cm	0.15	0.13-1.18
Toeslope, 5cm	0.15	0.14-1.12
Toeslope, 15cm	0.11	0.09-1.19
Toeslope, 30cm	0.09	0.08-1.20
Toeslope, 60cm	0.03	0.09-0.36
Observations	4	

Appendix C

C.1. Supplementary materials

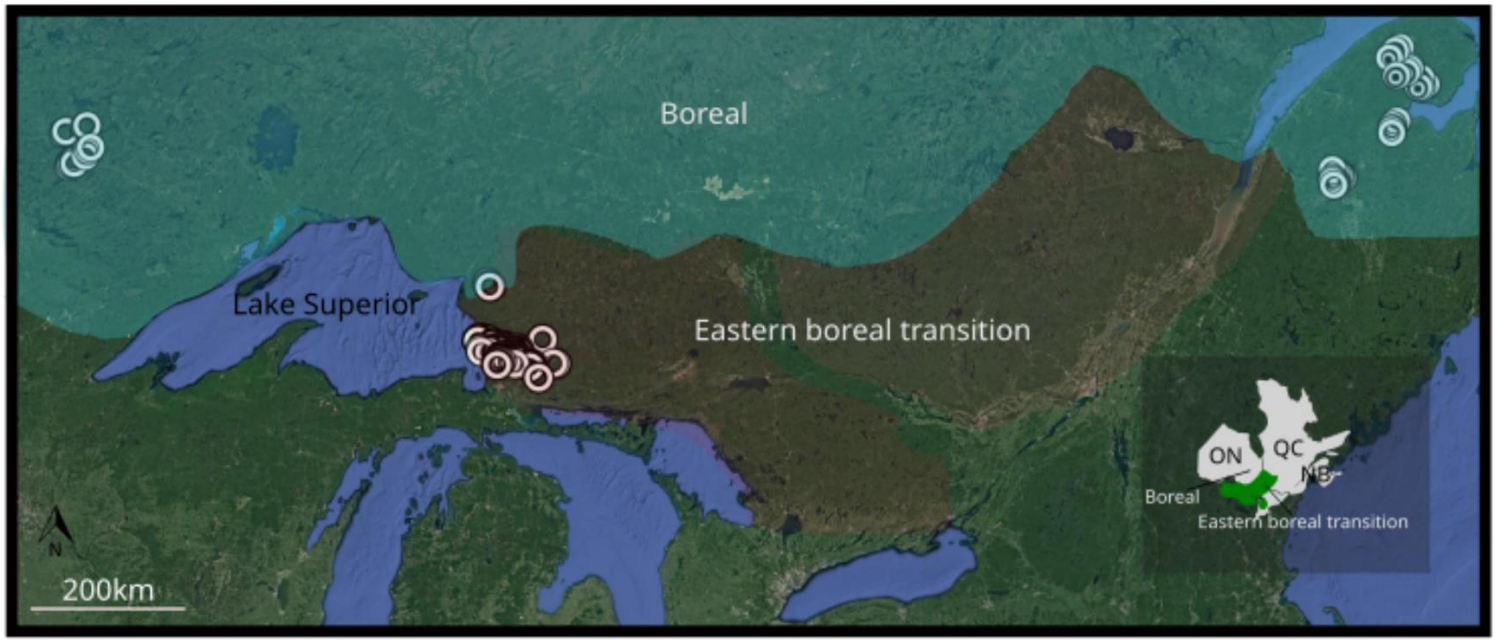


Fig. C.1 Streams sampled in Canada's eastern forest boreal transition and boreal forest. White circles represent sampling locations (n=268) where all samples were collected (n= 4589). Forest-type boundaries are drawn based on data contained in ref[420]. ON = Ontario, QC = Québec, and NB = New Brunswick are the Canadian provinces from which samples were collected.

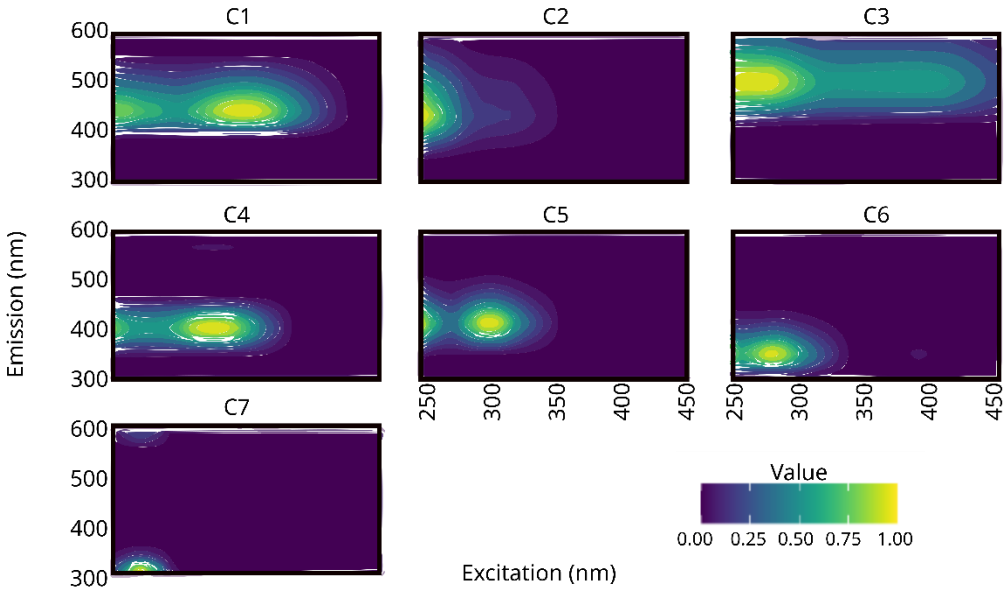


Fig. C.2 Excitation (x-axis) and emission (y-axis) contour plots of PARAFAC components.

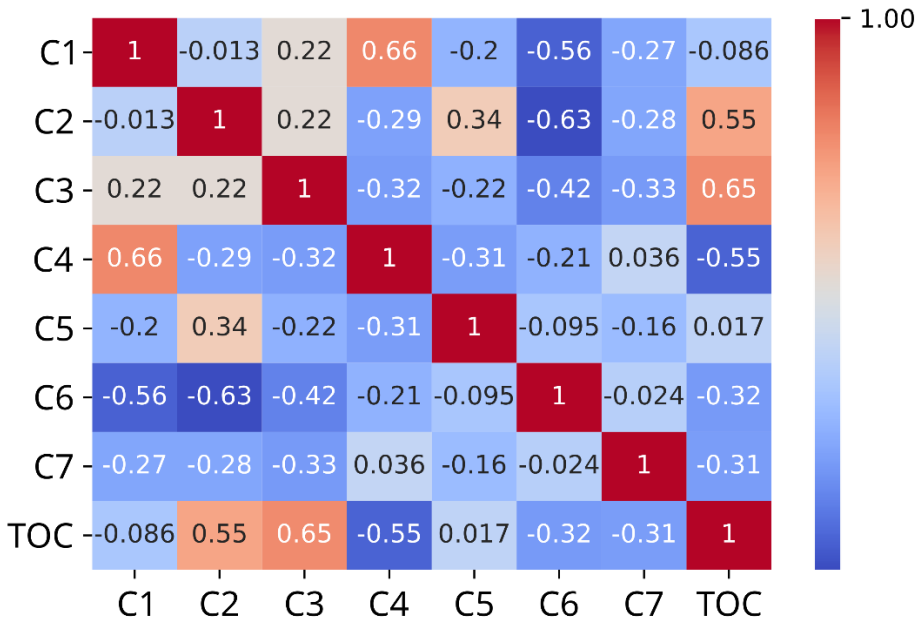


Fig. C.3 The correlation coefficients of relative compound raw intensities.

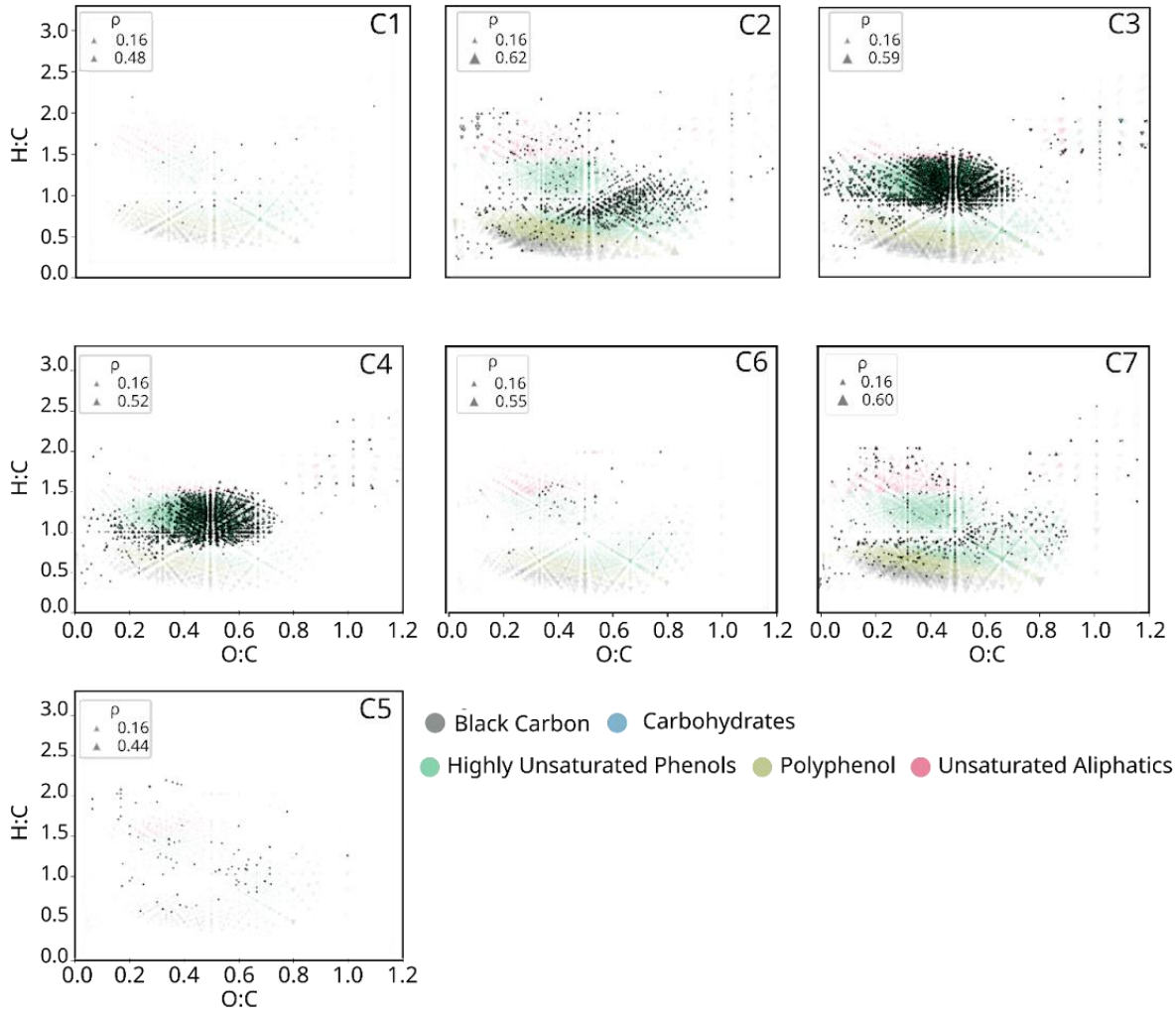


Fig. C.4 Compounds uniquely correlated (p<0.001) with PARAFAC components (C1-C7). Colours correspond to compound class. Upward triangles are positive correlations, and downward triangles are negative correlations. The size of the triangle is the strength of the correlation. Faded triangles are correlated compounds not unique to a single component. Mass spectral peak intensities were normalised to the sum intensity of all peaks within a sample. Pairwise Spearman's rank correlations (ρ) were then obtained between normalised PARAFAC components and normalised mass peak intensities. The classes are not mutually exclusive.

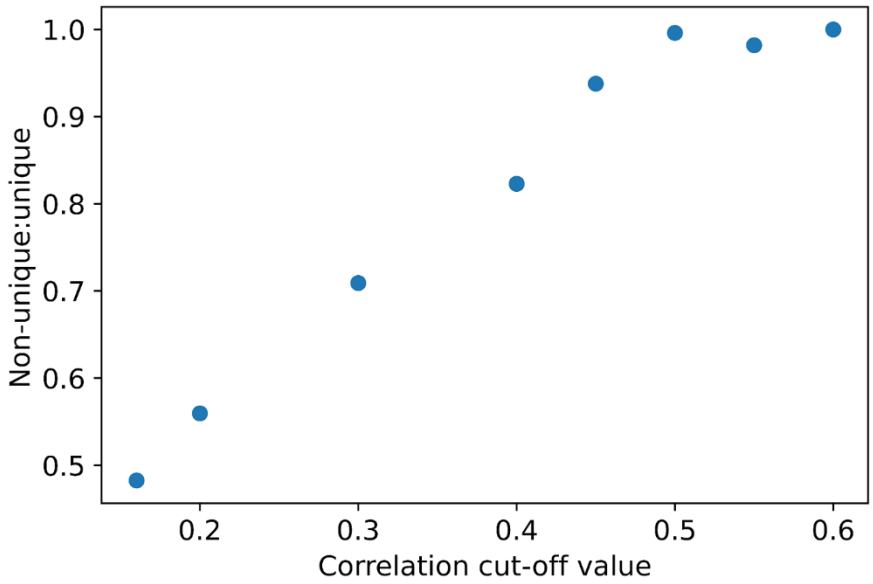


Fig. C.5 The basis for the correlation cut-off selection. A threshold of correlation strength between PARAFAC components and molecular formulae was selected based on the ratio of uniquely correlated compounds versus all compounds. The ratio plateaus between *a cut-off of* ρ= 0*.45-0.50.*

Table C.1 Excitation loading values for PARAFAC components.

Wavelength	C1	C2	C3	C4	C5	C6	C7
250	0.7906	1.0000	0.9808	0.7920	0.9303	0.9101	0.1995
255	0.8113	0.8925	0.9865	0.7046	0.7261	0.7009	0.4412
260	0.7981	0.7785	0.9931	0.6335	0.5959	0.6869	0.6782
265	0.7387	0.6881	0.9794	0.5835	0.4168	0.7519	0.9481
270	0.7129	0.5356	1.0000	0.5627	0.3849	0.8546	1.0000
275	0.6534	0.4324	0.9667	0.5364	0.4062	0.9429	0.8732
280	0.6724	0.2995	0.9740	0.5565	0.6050	1.0000	0.6670
285	0.6178	0.2347	0.9159	0.5651	0.7378	0.9527	0.4311
290	0.5813	0.1911	0.8585	0.6055	0.8636	0.8763	0.1572
295	0.5720	0.1615	0.8113	0.6644	0.9963	0.7427	0.0417
300	0.5601	0.1650	0.7402	0.7269	1.0000	0.6235	0.0110
305	0.6170	0.1516	0.6966	0.8308	0.9764	0.4985	0.0000
310	0.6730	0.1443	0.6472	0.9100	0.8617	0.3928	0.0000
315	0.7215	0.1482	0.5918	0.9556	0.6896	0.3064	0.0000
320	0.7937	0.1330	0.5664	0.9979	0.5268	0.2217	0.0000
325	0.8518	0.1139	0.5467	1.0000	0.4011	0.1514	0.0017
330	0.9076	0.0983	0.5417	0.9866	0.2998	0.0999	0.0064
335	0.9505	0.0787	0.5425	0.9461	0.2124	0.0503	0.0120
340	0.9830	0.0614	0.5449	0.8769	0.1392	0.0134	0.0166
345	0.9998	0.0468	0.5483	0.7790	0.0913	0.0000	0.0195
350	1.0000	0.0361	0.5534	0.6717	0.0396	0.0000	0.0179
355	0.9801	0.0150	0.5600	0.5478	0.0143	0.0000	0.0181
360	0.9333	0.0072	0.5595	0.4228	0.0000	0.0000	0.0168
365	0.8683	0.0118	0.5636	0.2992	0.0000	0.0000	0.0175
370	0.7996	0.0000	0.5758	0.2029	0.0000	0.0000	0.0173
375	0.7121	0.0000	0.5842	0.1205	0.0000	0.0113	0.0153
380	0.6189	0.0000	0.5834	0.0608	0.0000	0.0226	0.0128
385	0.5318	0.0000	0.5837	0.0169	0.0000	0.0320	0.0104
390	0.4393	0.0000	0.5809	0.0000	0.0000	0.0339	0.0100

Wavelength	C1	C2	C3	C4	C5	C6	C7
395	0.3541	0.0000	0.5754	0.0000	0.0000	0.0322	0.0093
400	0.2755	0.0000	0.5717	0.0000	0.0000	0.0283	0.0107
405	0.2070	0.0000	0.5543	0.0000	0.0000	0.0262	0.0104
410	0.1487	0.0000	0.5327	0.0000	0.0000	0.0271	0.0105
415	0.0989	0.0000	0.5060	0.0000	0.0000	0.0277	0.0103
420	0.0601	0.0000	0.4736	0.0000	0.0000	0.0281	0.0095
425	0.0316	0.0000	0.4359	0.0000	0.0000	0.0290	0.0101
430	0.0051	0.0000	0.4025	0.0034	0.0000	0.0293	0.0105
435	0.0000	0.0000	0.3590	0.0008	0.0000	0.0297	0.0110
440	0.0000	0.0000	0.3149	0.0000	0.0000	0.0291	0.0116
445	0.0000	0.0000	0.2767	0.0000	0.0000	0.0287	0.0120
450	0.0000	0.0000	0.2440	0.0000	0.0000	0.0283	0.0125

Table C.2 Emission loading values for PARAFAC components.

Wavelength	C1	C2	C3	C4	C5	C6	C7
300	0.000	0.000	0.021	0.000	0.033	0.105	0.878
302	0.000	0.000	0.025	0.000	0.026	0.098	0.922
304	0.000	0.000	0.027	0.003	0.008	0.101	0.950
306	0.000	0.000	0.030	0.003	0.000	0.098	0.986
308	0.000	0.000	0.030	0.001	0.000	0.108	1.000
310	0.000	0.005	0.026	0.000	0.000	0.130	0.989
312	0.000	0.003	0.025	0.000	0.000	0.160	0.974
314	0.000	0.000	0.022	0.000	0.000	0.197	0.937
316	0.000	0.000	0.019	0.000	0.000	0.245	0.910
318	0.000	0.000	0.014	0.000	0.000	0.307	0.856
320	0.000	0.000	0.010	0.000	0.000	0.382	0.812
322	0.000	0.000	0.002	0.000	0.000	0.451	0.723
324	0.000	0.000	0.000	0.000	0.000	0.514	0.622
326	0.000	0.000	0.000	0.000	0.000	0.564	0.547
328	0.000	0.000	0.000	0.000	0.000	0.609	0.489
330	0.000	0.000	0.000	0.000	0.000	0.660	0.404
332	0.000	0.000	0.000	0.000	0.000	0.713	0.352
334	0.000	0.000	0.000	0.000	0.000	0.770	0.300
336	0.000	0.000	0.000	0.000	0.000	0.835	0.263
338	0.000	0.000	0.000	0.000	0.000	0.878	0.221
340	0.000	0.000	0.007	0.000	0.000	0.912	0.175
342	0.000	0.001	0.011	0.010	0.000	0.936	0.162
344	0.000	0.007	0.009	0.026	0.000	0.934	0.130
346	0.000	0.015	0.007	0.041	0.004	0.948	0.109
348	0.000	0.025	0.005	0.057	0.018	0.984	0.107
350	0.000	0.034	0.002	0.073	0.041	0.984	0.087
352	0.000	0.043	0.000	0.095	0.056	1.000	0.079
354	0.000	0.055	0.000	0.115	0.080	0.994	0.060
356	0.000	0.060	0.000	0.150	0.088	0.982	0.050

Wavelength	C1	C2	C3	C4	C5	C6	C7
358	0.000	0.075	0.000	0.173	0.117	0.969	0.040
360	0.000	0.090	0.000	0.207	0.143	0.939	0.035
362	0.000	0.103	0.000	0.239	0.172	0.897	0.019
364	0.000	0.124	0.000	0.277	0.199	0.894	0.025
366	0.000	0.149	0.000	0.314	0.222	0.869	0.020
368	0.000	0.170	0.000	0.361	0.248	0.835	0.021
370	0.000	0.196	0.000	0.397	0.289	0.831	0.008
372	0.000	0.217	0.000	0.443	0.315	0.783	0.013
374	0.000	0.241	0.000	0.488	0.356	0.750	0.013
376	0.000	0.259	0.000	0.531	0.395	0.713	0.004
378	0.000	0.281	0.000	0.584	0.406	0.658	0.011
380	0.000	0.315	0.000	0.636	0.456	0.643	0.005
382	0.000	0.342	0.000	0.685	0.489	0.591	0.017
384	0.000	0.359	0.004	0.732	0.519	0.549	0.021
386	0.000	0.392	0.011	0.782	0.573	0.517	0.021
388	0.000	0.412	0.021	0.832	0.604	0.471	0.016
390	0.000	0.460	0.036	0.903	0.675	0.464	0.023
392	0.033	0.485	0.024	0.909	0.711	0.425	0.019
394	0.058	0.510	0.024	0.938	0.737	0.396	0.013
396	0.093	0.555	0.017	0.958	0.801	0.381	0.013
398	0.127	0.583	0.012	0.979	0.837	0.346	0.022
400	0.169	0.624	0.004	0.988	0.880	0.315	0.025
402	0.204	0.650	0.007	1.000	0.892	0.290	0.020
404	0.254	0.668	0.000	0.980	0.910	0.265	0.015
406	0.299	0.706	0.000	0.984	0.939	0.251	0.012
408	0.343	0.735	0.000	0.989	0.967	0.227	0.011
410	0.395	0.758	0.000	0.962	0.983	0.213	0.003
412	0.441	0.797	0.003	0.969	0.983	0.192	0.012
414	0.488	0.830	0.014	0.966	1.000	0.174	0.009

Wavelength	C1	C2	C3	C4	C5	C6	C7
416	0.531	0.852	0.030	0.954	0.994	0.153	0.011
418	0.578	0.879	0.041	0.929	0.980	0.130	0.017
420	0.634	0.896	0.043	0.873	0.966	0.121	0.015
422	0.674	0.897	0.058	0.827	0.940	0.111	0.007
424	0.726	0.921	0.070	0.787	0.935	0.104	0.000
426	0.773	0.941	0.090	0.749	0.927	0.093	0.000
428	0.814	0.964	0.111	0.716	0.903	0.081	0.000
430	0.849	0.978	0.143	0.688	0.873	0.071	0.000
432	0.876	0.984	0.177	0.659	0.840	0.059	0.006
434	0.908	0.990	0.196	0.609	0.815	0.052	0.004
436	0.935	1.000	0.227	0.567	0.785	0.049	0.002
438	0.948	0.995	0.269	0.531	0.743	0.039	0.009
440	0.969	0.996	0.296	0.487	0.719	0.039	0.000
442	0.984	0.992	0.327	0.434	0.679	0.036	0.000
444	0.991	0.996	0.366	0.419	0.647	0.027	0.001
446	0.996	0.977	0.396	0.354	0.598	0.025	0.006
448	1.000	0.974	0.423	0.306	0.583	0.024	0.000
450	1.000	0.959	0.457	0.277	0.549	0.017	0.000
452	0.994	0.954	0.493	0.250	0.509	0.017	0.005
454	0.976	0.933	0.519	0.215	0.483	0.013	0.003
456	0.960	0.910	0.552	0.180	0.441	0.012	0.000
458	0.930	0.883	0.581	0.160	0.397	0.003	0.007
460	0.946	0.901	0.609	0.132	0.393	0.006	0.000
462	0.935	0.886	0.630	0.104	0.378	0.007	0.000
464	0.904	0.863	0.663	0.094	0.345	0.000	0.000
466	0.876	0.842	0.689	0.072	0.310	0.000	0.001
468	0.851	0.820	0.705	0.052	0.292	0.002	0.000
470	0.831	0.802	0.728	0.036	0.269	0.000	0.000
472	0.821	0.803	0.753	0.023	0.258	0.000	0.000
474	0.794	0.785	0.771	0.010	0.239	0.000	0.000

Wavelength	C1	C2	C3	C4	C5	C6	C7
476	0.780	0.775	0.805	0.004	0.214	0.000	0.000
478	0.754	0.763	0.824	0.000	0.200	0.000	0.000
480	0.745	0.761	0.855	0.000	0.180	0.000	0.000
482	0.719	0.753	0.885	0.000	0.159	0.000	0.001
484	0.678	0.719	0.884	0.000	0.152	0.000	0.000
486	0.664	0.716	0.906	0.000	0.135	0.000	0.000
488	0.635	0.704	0.920	0.000	0.130	0.000	0.000
490	0.615	0.698	0.945	0.000	0.117	0.000	0.000
492	0.589	0.687	0.937	0.000	0.110	0.000	0.000
494	0.572	0.678	0.961	0.000	0.103	0.000	0.000
496	0.556	0.666	0.982	0.000	0.089	0.000	0.000
498	0.527	0.653	0.981	0.000	0.077	0.000	0.000
500	0.502	0.638	0.989	0.000	0.071	0.000	0.000
502	0.483	0.622	0.999	0.000	0.063	0.000	0.000
504	0.459	0.605	0.997	0.000	0.056	0.000	0.000
506	0.438	0.589	0.998	0.000	0.046	0.000	0.000
508	0.414	0.574	1.000	0.000	0.039	0.000	0.000
510	0.397	0.556	0.994	0.000	0.033	0.000	0.000
512	0.375	0.540	0.989	0.000	0.035	0.000	0.000
514	0.353	0.522	0.986	0.000	0.031	0.000	0.000
516	0.341	0.500	0.988	0.000	0.029	0.000	0.000
518	0.319	0.479	0.979	0.000	0.017	0.000	0.000
520	0.301	0.465	0.966	0.000	0.018	0.000	0.000
522	0.288	0.449	0.964	0.000	0.016	0.000	0.000
524	0.267	0.436	0.951	0.000	0.005	0.000	0.000
526	0.249	0.422	0.933	0.000	0.000	0.000	0.000
528	0.234	0.405	0.925	0.000	0.000	0.000	0.000
530	0.213	0.389	0.898	0.000	0.000	0.000	0.000
532	0.205	0.378	0.890	0.000	0.003	0.000	0.000

Wavelength	C1	C2	C3	C4	C5	C6	C7
534	0.186	0.358	0.865	0.000	0.002	0.000	0.000
536	0.175	0.341	0.860	0.000	0.000	0.000	0.000
538	0.161	0.334	0.833	0.000	0.002	0.000	0.000
540	0.152	0.322	0.815	0.000	0.000	0.000	0.000
542	0.133	0.309	0.787	0.001	0.000	0.000	0.000
544	0.120	0.296	0.774	0.001	0.000	0.000	0.000
546	0.108	0.280	0.763	0.008	0.000	0.000	0.000
548	0.099	0.270	0.747	0.008	0.000	0.000	0.000
550	0.089	0.257	0.736	0.011	0.000	0.000	0.000
552	0.078	0.246	0.716	0.016	0.000	0.000	0.000
554	0.068	0.236	0.698	0.017	0.000	0.000	0.003
556	0.059	0.223	0.684	0.021	0.000	0.000	0.007
558	0.050	0.212	0.658	0.024	0.000	0.000	0.014
560	0.038	0.198	0.649	0.027	0.000	0.000	0.021
562	0.033	0.189	0.627	0.028	0.000	0.000	0.032
564	0.025	0.177	0.606	0.030	0.000	0.000	0.041
566	0.015	0.165	0.591	0.036	0.000	0.000	0.052
568	0.009	0.153	0.574	0.036	0.000	0.000	0.063
570	0.004	0.139	0.571	0.041	0.000	0.000	0.081
572	0.000	0.133	0.546	0.039	0.000	0.000	0.097
574	0.000	0.125	0.524	0.033	0.000	0.000	0.112
576	0.000	0.119	0.496	0.031	0.000	0.000	0.124
578	0.000	0.111	0.477	0.027	0.000	0.000	0.139
580	0.000	0.101	0.458	0.027	0.000	0.000	0.157
582	0.000	0.100	0.431	0.022	0.000	0.000	0.161
584	0.000	0.093	0.413	0.021	0.000	0.000	0.176
586	0.000	0.088	0.394	0.021	0.000	0.000	0.185
588	0.000	0.083	0.376	0.020	0.000	0.000	0.194
590	0.000	0.079	0.356	0.017	0.000	0.000	0.203
592	0.000	0.073	0.343	0.014	0.000	0.000	0.207

Wavelength	C1	C2	C3	C4	C5	C6	C7
594	**0.000**	**0.070**	**0.328**	**0.010**	**0.000**	**0.000**	**0.214**
596	**0.000**	**0.065**	**0.313**	**0.013**	**0.000**	**0.000**	**0.211**
598	**0.000**	**0.063**	**0.302**	**0.010**	**0.000**	**0.000**	**0.211**

Table C.3 OpenFluor database matches with the seven PARAFAC components resolved by our model. We show excitation (Ex.) and emission (Em.) maxima of the seven PARAFAC components in (n=4589) followed by the component matches with the OpenFluor[419] database based on a Tucker's Congruence Coefficient (TCC) > 0.98. Match descriptions in terms of source and traditional peak name as well as the region of sample collection are given.

Component	Ex/Em	OpenFluor matching	Description	Zone
C1	**350/448**	**C1**[421]	Terrestrial humic-like	Aquaponic system
		C1[422]	Humic-like, traditional peak A+M	Soil samples from highly urbanized area of Beijing, China
		C1[423]	Terrestrially derived humic matter derived from lignin	Peat-draining rivers, estuaries and coastal waters of Sarawak, Borneo
		C2[424]	Terrestrial humic-like	The Meuse River, Belgium
		C3[425]	Peak C, terrestrial component identified in a wide range of environments	Danish streams and lakes
		C1[426]	Peak C, closely resembles microbial fulvic acids	27 prairie saline lake ecosystems
		C1[427]	UV-C humic like	Lake water
C2	**250/436**	**C3**[428]	Humic-like. Described as peak Af and related to terrestrially derived materials, associated with low molecular	4 boreal lakes
		C4[429]	humic-like	Northern lake ice
		C3[430]	Terrestrial humic-like	Phong River, Thailand Water Treatment Plant
		C3[423]	Humic/fulvic acid-like components derived from terrestrial plant litter	Tropical peatlands of coastal Sarawak, Borneo
		C2[431]	Terrestrial humic-like	Subtropical freshwater wetland, the Everglades, Florida, USA
		C1[437]	Peak A in the literature. Peak A fluorescence has been	Danish estuary, lakes, rivers

Component	Ex/Em	OpenFluor matching	Description	Zone
			observed in both marine and terrestrial CDOM	
		C1[432]	Terrestrial humic-like	Beaufort Sea and North Atlantic Subtropical Gyre
		C3[427]	Humic-like, high molecular weight aromatic	Six boreal lakes
		C1[433]	Associated with a terrestrial DOM source	Penobscot River, Androscoggin River, Penobscot Bay, and the Gulf of Maine
C3	**270/508**	**C3**[434]	Similar to another terrestrial humic-like component at the longest emission wavelength	Typical culture area of shellfish and algae in SE China
		C1[435]	No info	Adirondack Lake Watershed
		C2[436]	No info	Onondaga Lake–Three Rivers System
		C3[437]	Humic-like, terrestrial material	Stream sediments
		C2[438]	Indicator of allocthony	Subset of waters in subarctic Quebec
		C2[439]	Humic-like components	Stream in Hokkaido University's Uryu Experimental Forest (UREF), northern Hokkaido, Japan
		C2[440]	No info	111 lakes across New York
		C2[441]	Terrestrial humic-like	Rainwater, fresh plant, leaf litter, wastewater, and river from Minjiang Watershed, southeast China
		C3[442]	Humic-like, susceptible to loss by adsorption to clay	20 lakes, three peats and seven streams and rivers throughout Sweden, covering the regions of Jämtland, Bergslagen, Uppland, and Småland
		C3[421]	Humic-like intermediate of the photodegradation of humic compounds	Recirculation aquaponic systems

Component	Ex/Em	OpenFluor matching	Description	Zone
		C1[443]	Humic-like derived from terrestrial material by photochemical degradation	Tributaries lower Austria
		C2[444]	Terrestrial humic-like	Typical coastal area (Minjiang Estuary, China)
		C2[423]	Soil fulvic acid derived from terrestrial higher plants	Tropical peatlands of coastal Sarawak, Borneo
		C2[445]	Humic-like	(1) the Milwaukee River (2) open Green Bay (3) Veterans Lagoon
		C2[446]	Terrestrial humic	Neuse River Estuary, Charleston Harbor, and the inner Louisiana–Texas Shelf of the Gulf of Mexico
		C3[447]	Humic-like	A municipal drinking water distribution network in central eastern Sweden
		C4[448]	Humic-like, ubiquitous humic-like components	Antarctic lakes and streams
		C3[449]	Terrestrial humic-like	Land-ocean interface of a subtropical river (Minjiang River, SE China)
		C2[450]	Terrestrial humic-like, terrestrial reduced quinone-like	Tropical rivers of the Guayana Shield, Venezuela
		C3[451]	Terrestrial humic-like	Coastal Canadian Arctic surface waters
		C4[425]	Similar to soil fulvic peak. High molecular weight material	Different Danish freshwaters
		C2[452]	Derivatives of terrestrial organic matter	Sea ice in coastal waters of the Baltic Sea
		C2[453]	Terrestrial signal, intermediate formed during photochemical degradation of terrestrial organic matter	N/A
		C3[424]	Humic-like	Ballast Water
		C2[454]	Terrestrial humic-like, high molecular weight	Along the Zambezi River

Component	Ex/Em	OpenFluor matching	Description	Zone
		C2[455]	Humic-like, terrestrial	Central European headwater streams
C4	**325/402**	**C1**[456]	Humic-like	Ter River–Sau Reservoir system (Spain)
		C2[457]	Humic-like, terrestrial and marine origin	Submarine mud volcanoes in the Gulf of Cádiz
		C2[458]	Humic-like	Groundwater sampled in the Bengal Basin
		C2[459]	Terrestrial humic-like	Holocene and Pleistocene aquifers
		C2[460]	No info	Eutrophic estuary
		C6[431]	Ubiquitous humic-like	Subtropical freshwater wetland, the Everglades, Florida, USA.
		C3[461]	Marine humic-substances	Coastal South Atlantic Bight
		C2[462]	Humic-like freshwater lakes, saline lakes, rivers, and estuaries	Recirculating aquaculture systems
C5	**300/414**	**C4**[428]	Humic-like materials likely the result of degradation process	Lake water was collected from 4 different lakes in east-central Sweden
		C5[463]	Humic-like components, resembling refractory types of organic matter but also generated in situ during microbial degradation	Sea Ice in the Canada Basin, Arctic Ocean
		C4[443]	Artifact from the fluorometer, has no ecologic implication	Streams in Western part of Lower Austria
C6	**280/352**	**C3**[464]	Non-humic associated with alipahtic compounds and low-molecular weight, protein-like	Temperate lakes of Southern Andes (Patagonia, Argentina)
		C4[465]	No info	Three Fjords: Kongsfjorden (Longyearbyen), Scoresby Sound (East Greenland), and Arnarfjörður (West Iceland)

Component	Ex/Em	OpenFluor matching	Description	Zone
		C6[466]	Tryptophan-like, aliphatic, low molecular weight	Congo River network
		C5[467]	Tryptophan-like, protein-like	Headwater Streams, US
		C5[468]	Terrestrial humic-like	Baltic–North Sea transition zone
		C7[469]	Tryptophan-like fluorescence derived from autocthonous	Temperate estuary, Denmark
		C4[470]	Peak T + Peak B, Protein-like, Microbial delivered	Water treatment plants, Queensland, Australia
		C5[433]	Protein-like component similar to the amino acid tryptophan	Penobscot River, Androscoggin River, Penobscot Bay, and the Gulf of Maine
		C4[471]	Protein-like	Shark Bay, Australia
C7	**270/308**	**C5**[472]	Tyrosine-like	Three coastal drainage basins of Miami, FL
		C3[473]	Protein-like, tyrosine-like	Chukchi Sea of the Arctic Ocean
		C1[474]	Tyrosine-like, protein-like	Arctic fjord
		C2[475]	Tyrosine-like, freshly produced proteinaceous material	Third order stream located in southwestern Montana, USA
		C2[476]	Tyrosine and Tryptophan, amino acids, free or bound in protein compounds	HLGW, upper Heihe Basin, Northeastern QTP, China

Table C.4 Summary of the counts of all molecules correlated with PARAFAC components and the features of their correlations. Where correlation is defined at the 99% confidence limit (and not by ρ cut-off).

Category	***Count***
Total number of molecules	**7412**
Total number of molecules that UNIQUELY correlate positively with at least one component	**2014**
Total number of molecules that UNIQUELY correlate negatively with at least one component	**1809**
Total number of molecules that UNIQUELY correlate with at least one component	**2737**
Total number of molecules that correlate with at least one component	**5673**
Total number of correlated molecules with N	**1669**
Total number of correlated molecules with S	**483**
Total number of correlated molecules with P	**336**

Table C.5 Summary of the pairwise PARAFAC components intersection, union, and proportion of strongly correlated molecules. Correlation is defined at ρ >|0.45|. (A) The number of intersecting correlated molecules shared by components, (B) the number of correlated compounds in the union of two components, (C) the proportion of overlapping correlated molecules between two components. Union is the set of all elements in either set or both, whereas intersection is the set of all distinct elements that belong to both sets.

A.

Intersection	***C2***	***C3***	***C4***	***C5***	***C6***	***C7***
C1	**8**	**7**	**0**	**0**	**6**	**8**
C2		**179**	**10**	**0**	**22**	**275**
C3			**27**	**0**	**23**	**193**
C4				**0**	**0**	**11**
C5					**0**	**0**
C6						**22**

B.

Union	***C2***	***C3***	***C4***	***C5***	***C6***	***C7***
C1	**328**	**257**	**82**	**8**	**28**	**320**
C2		**405**	**392**	**328**	**332**	**373**
C3			**303**	**256**	**259**	**383**
C4				**74**	**100**	**383**
C5					**26**	**320**
C6						**324**

C.

Proportion Overlap	***C2***	***C3***	***C4***	***C5***	***C6***	***C7***
C1	**0.02**	**0.03**	**0.00**	**0.00**	**0.21**	**0.03**
C2		**0.44**	**0.03**	**0.00**	**0.07**	**0.74**
C3			**0.09**	**0.00**	**0.09**	**0.50**
C4				**0.00**	**0.00**	**0.03**
C5					**0.00**	**0.00**
C6						**0.07**

Table C.6 Summary of the pairwise PARAFAC components intersection, union, and proportion of all correlated molecules.

Correlation is defined at the 99% confidence limit. (A) The number of intersecting correlated molecules shared by components, (B) the number of correlated compounds in the union of two components, (C) the proportion of overlapping correlated molecules between two components. Union is the set of all elements in either set or both, whereas intersection is the set of all distinct elements that belong to both sets.

A.

Intersection	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*
C1	**1748**	**1431**	**1178**	**909**	**1385**	**1738**
C2		**2702**	**2404**	**1376**	**2066**	**3348**
C3			**3509**	**1101**	**1739**	**2966**
C4				**953**	**1467**	**2650**
C5					**1191**	**1367**
C6						**2156**

B.

Union	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*
C1	**3797**	**4460**	**4545**	**2408**	**2678**	**3917**
C2		**5126**	**5256**	**3878**	**3934**	**4244**
C3			**4497**	**4499**	**4607**	**4972**
C4				**4479**	**4711**	**5120**
C5					**2581**	**3997**
C6						**3954**

C.

Proportion Overlap	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*
C1	**0.46**	**0.32**	**0.26**	**0.38**	**0.52**	**0.44**
C2		**0.53**	**0.46**	**0.35**	**0.53**	**0.79**
C3			**0.78**	**0.24**	**0.38**	**0.60**
C4				**0.21**	**0.31**	**0.52**
C5					**0.46**	**0.34**
C6						**0.55**

Table C.7 ***Summary of counts of molecules strongly correlated and the features of their strong correlations.***

Correlation is defined as ρ >0.45: the count of the compound classes that are uniquely correlated with a component, correlated with a component (1 or more), and the total count of that class in the dataset (n=7412), the classes are not mutually exclusive.

Category	*Count*
Total Number of Molecules	***7412***
Total number of molecules that UNIQUELY correlate positively with at least one component:	***237***
Total number of molecules that UNIQUELY correlate negatively with at least one component	***342***
Total number of molecules that UNIQUELY correlate with at least one component	***452***
Total number of molecules that correlate with at least one component	***482***
Total number of correlated molecules with N	***93***
Total number of correlated molecules with S	***10***
Total number of correlated molecules with P	***4***

Table C.8 Summary of compounds strongly correlated with PARARAC components by compound class. Correlation is defined at the ρ >|0.45|. The classes are not mutually exclusive. BC = Black Carbon, BC_C15min = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C<15. BC_C15max >C14 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C >= 15. Polyphen= Polyphenols, Polyphen_Orich = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol_Opoor= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUPh=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations.

	Uniquely Correlated	***Correlated***	***Total***
BC	**157**	**72**	**167**
BC_C15min	**88**	**39**	**110**
BC_C15max	**69**	**33**	**57**
BC_X	**5**	**3**	**178**
Polyphen	**384**	**179**	**766**
Polyphen_Orich	**81**	**41**	**90**
Polyphen_Opoor	**303**	**138**	**676**
HUPh	**326**	**216**	**5318**
HUP_Orich	**132**	**66**	**2306**
HUP_Opoor	**194**	**150**	**3012**
UnsatAliph	**10**	**8**	**565**
UnsatAliph_Orich	**2**	**2**	**150**

	Uniquely Correlated	*Correlated*	*Total*
UnsatAliph_Opoor	**8**	**6**	**415**
SatFA_CHOX	**0**	**0**	**66**
Carbo_CHO	**4**	**2**	**27**
Carbo_CHOX	**2**	**2**	**224**
Pep	**0**	**0**	**108**
CRAM	**185**	**146**	**3560**
CHO	**885**	**480**	**7311**
CHOP	**4**	**4**	**530**
CHOS	**9**	**8**	**725**
CHON	**115**	**91**	**2368**
CHOSP	**0**	**0**	**42**
CHONP	**0**	**0**	**3**
CHONS	**2**	**1**	**24**

Table C.9 For each PARAFAC component a summary of all positively correlated compounds in a compound class as a proportion of all correlated compounds in that compound class. Where the correlation was defined as at the 99% confidence limit. Correlations are with at least one component. The classes are not mutually exclusive. BC = Black Carbon, BC_C15min = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C<15. BC_C15max >C14 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C >= 15. Polyphen= Polyphenols, Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUP=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations.

	C1	C2	C3	C4	C5	C6	C7
BC	**0.01**	**0.32**	**0.28**	**0.00**	**0.00**	**0.00**	**0.00**
BC_C15min	**0.00**	**0.33**	**0.26**	**0.00**	**0.00**	**0.00**	**0.00**
BC_C15max	**0.01**	**0.32**	**0.30**	**0.00**	**0.00**	**0.00**	**0.00**
BC_X	**0.00**	**0.20**	**0.00**	**0.00**	**0.00**	**0.20**	**0.00**
Polyphen	**0.00**	**0.38**	**0.24**	**0.00**	**0.00**	**0.00**	**0.00**
Polyphen_Orich	**0.01**	**0.28**	**0.36**	**0.00**	**0.00**	**0.00**	**0.00**
Polyphen_Opoor	**0.00**	**0.41**	**0.20**	**0.00**	**0.00**	**0.00**	**0.00**
HUP	**0.00**	**0.27**	**0.08**	**0.20**	**0.00**	**0.01**	**0.04**
HUP_Orich	**0.00**	**0.43**	**0.19**	**0.03**	**0.00**	**0.00**	**0.00**
HUP_Opoor	**0.00**	**0.16**	**0.01**	**0.32**	**0.00**	**0.02**	**0.07**
UnsatAliph	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.60**
UnsatAliph_Orich	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
UnsatAliph_Opoor	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.75**
SatFA_CHOX	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Carbo_CHO	**0.00**	**0.50**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Carbo_CHOX	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

	C1	C2	C3	C4	C5	C6	C7
Pep	-	-	-	-	-	-	-
CRAM	0.00	0.15	0.01	0.35	0.00	0.02	0.06
CHO	0.00	0.33	0.18	0.07	0.00	0.00	0.02
CHOP	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHOS	0.00	0.00	0.00	0.00	0.00	0.44	0.00
CHON	0.00	0.00	0.00	0.57	0.00	0.01	0.09
CHOSP	-	-	-	-	-	-	-
CHONP	-	-	-	-	-	-	-
CHONS	0.00	0.00	0.00	0.00	0.00	0.50	0.00

Table C.10 For each PARAFAC component a summary of all strongly positively correlated compounds in a compound class as a proportion of all correlated compounds in that compound class. The correlation was defined as at ρ >0.45. The correlation was defined as at the 99% confidence limit. The classes are not mutually exclusive. BC = Black Carbon, BC_C15min = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C<15. BC_C15max >C14 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C >= 15. Polyphen= Polyphenols, Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUP=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations.

	C1	C2	C3	C4	C5	C6	C7
BC	0.01	0.32	0.28	0.00	0.00	0.00	0.00
BC_C15min	0.00	0.33	0.26	0.00	0.00	0.00	0.00
BC_C15max	0.01	0.32	0.30	0.00	0.00	0.00	0.00
BC_X	0.00	0.20	0.00	0.00	0.00	0.20	0.00
Polyphen	0.00	0.38	0.24	0.00	0.00	0.00	0.00
Polyphen_Orich	0.01	0.28	0.36	0.00	0.00	0.00	0.00
Polyphen_Opoor	0.00	0.41	0.20	0.00	0.00	0.00	0.00
HUPh	0.00	0.27	0.08	0.20	0.00	0.01	0.04
HUP_Orich	0.00	0.43	0.19	0.03	0.00	0.00	0.00
HUP_Opoor	0.00	0.16	0.01	0.32	0.00	0.02	0.07
UnsatAliph	0.00	0.00	0.00	0.00	0.00	0.00	0.60
UnsatAliph_Orich	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UnsatAliph_Opoor	0.00	0.00	0.00	0.00	0.00	0.00	0.75
SatFA_CHOX	-	-	-	-	-	-	-
Carbo_CHO	0.00	0.50	0.00	0.00	0.00	0.00	0.00

	C1	C2	C3	C4	C5	C6	C7
Carbo_CHOX	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pep	-	-	-	-	-	-	-
CRAM	0.00	0.15	0.01	0.35	0.00	0.02	0.06
CHO	0.00	0.33	0.18	0.07	0.00	0.00	0.02
CHOP	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHOS	0.00	0.00	0.00	0.00	0.00	0.44	0.00
CHON	0.00	0.00	0.00	0.57	0.00	0.01	0.09
CHOSP	-	-	-	-	-	-	-
CHONP	-	-	-	-	-	-	-
CHONS	0.00	0.00	0.00	0.00	0.00	0.50	0.00

Table C.11 For each PARAFAC component a summary of all negatively correlated compounds in a compound class as a proportion of all correlated compounds in that compound class. The correlation was defined at the 99% confidence limit. The classes are not mutually exclusive. BC = Black Carbon, BC_C15min = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C<15. BC_C15max >C14 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C >= 15. Polyphen= Polyphenols, Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUP=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations. Polyphen= Polyphenols, Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUP=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations.

	C1	C2	C3	C4	C5	C6	C7
BC	0.00	0.00	0.00	0.00	0.00	0.00	0.37
BC_C15min	0.00	0.00	0.00	0.00	0.00	0.00	0.33
BC_C15max	0.00	0.00	0.00	0.00	0.00	0.00	0.44
BC_X	0.00	0.00	0.01	0.00	0.00	0.00	0.01
Polyphen	0.00	0.00	0.00	0.00	0.00	0.00	0.19
Polyphen_Orich	0.00	0.00	0.00	0.00	0.00	0.00	0.31

	C1	C2	C3	C4	C5	C6	C7
Polyphen_Opoor	0.00	0.00	0.00	0.00	0.00	0.00	0.17
HUPh	0.00	0.00	0.01	0.00	0.00	0.00	0.01
HUP_Orich	0.00	0.00	0.00	0.00	0.00	0.00	0.02
HUP_Opoor	0.00	0.00	0.02	0.00	0.00	0.00	0.01
UnsatAliph	0.00	0.00	0.00	0.00	0.00	0.00	0.00
UnsatAliph_Orich	0.00	0.00	0.01	0.00	0.00	0.00	0.00
UnsatAliph_Opoor	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SatFA_CHOX	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Carbo_CHO	0.00	0.00	0.00	0.00	0.00	0.00	0.07
Carbo_CHOX	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Pep	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CRAM	0.00	0.00	0.02	0.00	0.00	0.00	0.00
CHO	0.00	0.00	0.01	0.00	0.00	0.00	0.04
CHOP	0.00	0.00	0.01	0.00	0.00	0.00	0.00
CHOS	0.00	0.00	0.01	0.00	0.00	0.00	0.00
CHON	0.00	0.00	0.01	0.00	0.00	0.00	0.00
CHOSP	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHONP	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHONS	0.00	0.00	0.04	0.00	0.00	0.00	0.00

Table C.12 For each PARAFAC component a summary of all strongly negatively correlated compounds in a compound class as a proportion of all correlated compounds in that compound class.

The correlation was defined as at ρ >0.45. The classes are not mutually exclusive. BC = Black Carbon, BC_C15min = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C<15. BC_C15max >C14 = combustion-derived polycyclic aromates (PCAs, black carbon) without N, S, or P and C >= 15. Polyphen= Polyphenols, Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUP=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations. Polyphen= Polyphenols, Polyphenol high oxygen = O-rich soil-derived polyphenols and PCAs with aliphatic chains. Polyphenol low oxygen= O-poor soil-derived polyphenols and PCAs with aliphatic chains. HUP=Highly unsaturated phenols, HUP_Orich = O-rich highly unsaturated compounds, soil-derived humics, phenolics. HUP_Opoor =O-poor highly unsaturated compounds, soil-derived humics, phenolics. UnsatAliph =Unsaturated aliphatics, UnsatAliph_Orich = O-rich unsaturated aliphatic compounds, UnsatAliph_Opoor= O-poor unsaturated aliphatic compounds, SatFA_CHOX= saturated fatty acids CHOX, with N, S or P. CarboCHO = Carbohydrates, sugars without N, S, or P. Carbo_CHOX = carbohydrates, sugars with N, S, or P. Pep= peptides (unsaturated aliphatic and with at least 1 N). CRAM = Carboxyl-rich alicyclic molecules. The remaining rows correspond to elemental combinations.

	C1	C2	C3	C4	C5	C6	C7
BC	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.39**
BC_C15min	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.41**
BC_C15max	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.36**
BC_X	**0.00**	**0.00**	**0.20**	**0.00**	**0.00**	**0.00**	**0.40**
Polyphen	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	**0.00**	**0.37**

	C1	C2	C3	C4	C5	C6	C7
Polyphen_Orich	0.00	0.00	0.00	0.00	0.00	0.00	0.35
Polyphen_Opoor	0.00	0.00	0.01	0.00	0.00	0.00	0.38
HUPh	0.00	0.02	0.17	0.00	0.00	0.00	0.19
HUP_Orich	0.00	0.00	0.01	0.00	0.00	0.00	0.34
HUP_Opoor	0.00	0.04	0.29	0.00	0.00	0.00	0.09
UnsatAliph	0.00	0.20	0.20	0.00	0.00	0.00	0.00
UnsatAliph_Orich	0.00	0.00	1.00	0.00	0.00	0.00	0.00
UnsatAliph_Opoor	0.00	0.25	0.00	0.00	0.00	0.00	0.00
SatFA_CHOX	-	-	-	-	-	-	-
Carbo_CHO	0.00	0.00	0.00	0.00	0.00	0.00	0.50
Carbo_CHOX	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Pep	-	-	-	-	-	-	-
CRAM	0.00	0.04	0.30	0.00	0.00	0.00	0.08
CHO	0.00	0.01	0.07	0.00	0.00	0.00	0.30
CHOP	0.00	0.00	1.00	0.00	0.00	0.00	0.00
CHOS	0.00	0.00	0.56	0.00	0.00	0.00	0.00
CHON	0.00	0.06	0.27	0.00	0.00	0.00	0.00
CHOSP	-	-	-	-	-	-	-
CHONP	-	-	-	-	-	-	-
CHONS	0.00	0.00	0.50	0.00	0.00	0.00	0.00

Table C.13 Landscape characteristics used to predict PARAFAC components.
Ecotype descriptions are coded starting with an ecosite number, followed by a vegetative modifier, substrate depth modifier, substrate moisture modifier, substrate chemistry modifier, and vegetative cover class modifier. Details can be found in the Ontario forest resource inventory manual (page 68).

Topography	Tree Type (% cover)	Tree Type Scientific Name	Ecotype (% cover)	Ecotype (% cover) continued	Land Class (%) Cover	Harvest (n=348)
Median Sedimentation Index	Ash, black	*Fraxinus nigra*	146S	107TtM	Water	Harvest Area
Median Height Above Network Drainage	Balsam Fir	*Abies balsamea*	070TtD	107TtD	Forest	lin 1-99
Std Height Above Network Drainage	Birch, white (or paper)	*Betula papyrifera*	098TtD	104TtM	Large wetlands	inp 0.5-1.5
Area	Black Spruce	*Picea mariana*	108TtM	104TtD	Treed Wetland	step 1-99
Majority Aspect	Hemlock, eastern	*Tsuga Canadensis*	099TtD	101TtD	Unclassified	ex 0-0.9
Std Probability of Depression	Iron wood	*Ostrya virginiana*	042TtM	100TtD	Open Wetland	
Std Wetness	Jack Pine	*Pinus banksiana*	223TtD	088TtMD	Brush and Alder	
Std Sedimentation Index	Oak, red	*Quercus rubra*	224TtM	074TtM		
Median Slope	Red Maple	*Acer rubrum*	139S D	074TtD		
Median Wetness	Red Pine	*Pinus resinosa*	197X	073TtD		
Std Slope	Sugar Maple	*Acer saccharum*	042TtD	070TlD		
	Tamarack	*Larix laricina*	050TtM	069TtD		

Supplementary Methods

Time-weighting functions:

The first model (eq C.1) is known as the step decay function, and is commonly used in cumulative measures.

$$f(t_{ij}) = \begin{cases} 1 & \text{for } t_{ij} \leq T \\ 0 & \text{for } t_{ij} > T \end{cases} \quad \textbf{(C.1)}$$

Where tij is the weighted loss between time *i* and time *j*, and T is the time cut-off in years.

The second model (eq. C.2) is a linear decay. Weights decay linearly until the time cost cutoff is reached. From this point onward weights assume the value of 0.

$$f(t_{ij}) = \begin{cases} (1 - t_{ij}/T) & \text{for } t_{ij} \leq T \\ 0 & \text{for } t_{ij} > T \end{cases} \quad \textbf{(C.2)}$$

Where tij is the weighted loss between time *i* and time *j*, and T is the time cut-off in years.

The third model (eq. 3) is a negative exponential, with the following form:

$$f(t_{ij}) = e^{(-\beta t_{ij})} \quad \textbf{(C.3)}$$

Where tij is the weighted loss between time i and time j, and β is the parameter defining speed of decay.

Finally, we modelled the time weights using an inverse power decay.

$$f(t_{ij}) = \begin{cases} 1 & \text{for } t_{ij} \leq 1 \\ t_{ij}^{-\beta} & \text{for } t_{ij} > 1 \end{cases} \quad \textbf{(C.4)}$$

Where tij is the weighted loss between time i and time j, and β is the parameter defining speed of decay

References

1. Liu, N. *et al.* Forested lands dominate drinking water supply in the conterminous United States. *Environ. Res. Lett.* **16**, 084008 (2021).
2. Springgay, E., Casallas Ramirez, S., Janzen, S. & Vannozzi Brito, V. The Forest–Water Nexus: An International Perspective. *For. Trees Livelihoods* **10**, 915 (2019).
3. Andréassian, V. Waters and forests: from historical controversy to scientific debate. *J. Hydrol.* **291**, 1–27 (2004).
4. Farley, K. A., Jobbagy, E. G. & Jackson, R. B. Effects of afforestation on water yield: a global synthesis with implications for policy. *Glob. Chang. Biol.* **11**, 1565–1576 (2005).
5. Ellison, D., Futter, M. N. & Bishop, K. On the forest cover–water yield debate: from demand- to supply-side thinking. *Global Change Biology* vol. 18 806–820 Preprint at https://doi.org/10.1111/j.1365-2486.2011.02589.x (2012).
6. Ellison, D. *et al.* Trees, forests and water: Cool insights for a hot world. *Glob. Environ. Change* **43**, 51–61 (2017).
7. Zhang, M. & Wei, X. Deforestation, forestation, and water supply. *Science* **371**, 990–991 (2021).
8. Lewis, S. L. The Paris Agreement has solved a troubling problem. *Nature* **532**, 283 (2016).
9. Griscom, B. W. *et al.* Natural climate solutions. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11645–11650 (2017).
10. Sohi, S. P. Agriculture. Carbon storage with benefits. *Science* **338**, 1034–1035 (2012).
11. Fargione, J. E. *et al.* Natural climate solutions for the United States. *Sci Adv* **4**, eaat1869 (2018).
12. Heck, V., Gerten, D., Lucht, W. & Popp, A. Biomass-based negative emissions difficult to reconcile with planetary boundaries. *Nat. Clim. Chang.* **8**, 151–155 (2018).

13. Friedlingstein, P., Allen, M., Canadell, J. G., Peters, G. P. & Seneviratne, S. I. Comment on “The global tree restoration potential.” *Science* **366**, eaay8060 (2019).
14. Grainger, A., Iverson, L. R., Marland, G. H. & Prasad, A. Comment on “The global tree restoration potential.” *Science* **366**, eaay8334 (2019).
15. Lewis, S. L., Mitchard, E. T. A., Prentice, C., Maslin, M. & Poulter, B. Comment on “The global tree restoration potential.” *Science* **366**, eaaz0388 (2019).
16. Scharlemann, J. P. W., Tanner, E. V. J., Hiederer, R. & Kapos, V. Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Management* **5**, 81–91 (2014).
17. Cambi, M., Certini, G., Neri, F. & Marchi, E. The impact of heavy traffic on forest soils: A review. *For. Ecol. Manage.* **338**, 124–138 (2015).
18. Shah, N. W. *et al.* The effects of forest management on water quality. *For. Ecol. Manage.* **522**, 120397 (2022).
19. Schelker, J., EklöF, K., Bishop, K. & Laudon, H. Effects of forestry operations on dissolved organic carbon concentrations and export in boreal first-order streams. *J. Geophys. Res. D: Atmos.* **117**, 1011 (2012).
20. Battin, T. J. *et al.* The boundless carbon cycle. *Nat. Geosci.* **2**, 598–600 (2009).
21. Manzoni, S., Taylor, P., Richter, A., Porporato, A. & Ågren, G. I. Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. *New Phytol.* **196**, 79–91 (2012).
22. Butman, D. *et al.* Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 58–63 (2016).
23. Drake, T. W., Raymond, P. A. & Spencer, R. G. M. Terrestrial carbon inputs to inland waters: A current synthesis of estimates and uncertainty. *Limnol. Oceanogr. Lett.* **3**, 132–142 (2018).
24. Poorter, L. *et al.* Biomass resilience of Neotropical secondary forests. *Nature* **530**, 211–214 (2016).
25. Fuss, S. *et al.* Betting on negative emissions. *Nat. Clim. Chang.* **4**, 850–853 (2014).

26. Denis, M. *et al.* A comparative study on the pore-size and filter type effect on the molecular composition of soil and stream dissolved organic matter. *Org. Geochem.* **110**, 36–44 (2017).
27. Simpson, A. J., Simpson, M. J. & Soong, R. Nuclear Magnetic Resonance Spectroscopy and Its Key Role in Environmental Research. *Environmental Science & Technology* vol. 46 11488–11496 Preprint at https://doi.org/10.1021/es302154w (2012).
28. Zark, M. & Dittmar, T. Universal molecular structures in natural dissolved organic matter. *Nat. Commun.* **9**, 3178 (2018).
29. Aitkenhead-Peterson, J. A., McDowell, W. H. & Neff, J. C. Sources, Production, and Regulation of Allochthonous Dissolved Organic Matter Inputs to Surface Waters. *Aquatic Ecosystems* 25–70 Preprint at https://doi.org/10.1016/b978-012256371-3/50003-2 (2003).
30. Catalán, N., Marcé, R., Kothawala, D. N. & Tranvik, L. J. Organic carbon decomposition rates controlled by water retention time across inland waters. *Nat. Geosci.* **9**, 501–504 (2016).
31. Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M. & Hess, L. L. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO2. *Nature* **416**, 617–620 (2002).
32. Post, W. M. Organic Carbon in Soil and the Global Carbon Cycle. in *The Global Carbon Cycle* 277–302 (Springer Berlin Heidelberg, 1993).
33. Eswaran, H., Van Den Berg, E. & Reich, P. Organic carbon in soils of the world. *Soil Sci. Soc. Am. J.* **57**, 192–194 (1993).
34. Engel, M. H. & Macko, S. A. *Organic Geochemistry: Principles and Applications*. (Springer Science & Business Media, 2013).
35. Huc, A. Y. Origin and formation of organic matter in recent sediments and its relation to kerogen. *Kerogen* (1980).
36. Goñi, M. A. & Hedges, J. I. Potential applications of cutin-derived CuO reaction products for discriminating vascular plant sources in natural environments. *Geochim. Cosmochim. Acta* **54**, 3073–3081 (1990).
37. Weinhold, A. *et al.* Tree species richness differentially affects the chemical composition of leaves, roots and root exudates in four subtropical tree species. *J. Ecol.* **110**, 97–116 (2022).

38. Defossez, E. *et al.* Spatial and evolutionary predictability of phytochemical diversity. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
39. Walker, T. W. N. *et al.* Functional Traits 2.0: The power of the metabolome for ecology. *J. Ecol.* **110**, 4–20 (2022).
40. Kaiser, K. & Kalbitz, K. Cycling downwards – dissolved organic matter in soils. *Soil Biol. Biochem.* **52**, 29–32 (2012).
41. Kallenbach, C. M., Frey, S. D. & Grandy, A. S. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nat. Commun.* **7**, 13630 (2016).
42. Clemmensen, K. E. *et al.* Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science* **339**, 1615–1618 (2013).
43. Liang, C., Schimel, J. P. & Jastrow, J. D. The importance of anabolism in microbial control over soil carbon storage. *Nature Microbiology* **2**, 1–6 (2017).
44. Qualls, R. G. & Haines, B. L. Biodegradability of dissolved organic matter in forest throughfall, soil solution, and stream water. *Soil Sci. Soc. Am. J.* **56**, 578–586 (1992).
45. Kaiser, K., Canedo-Oropeza, M., McMahon, R. & Amon, R. M. W. Origins and transformations of dissolved organic matter in large Arctic rivers. *Sci. Rep.* **7**, 13064 (2017).
46. Kögel-Knabner, I. & Amelung, W. Dynamics, chemistry, and preservation of organic matter in soils. in *Treatise on Geochemistry* 157–215 (Elsevier, 2014).
47. Roth, V.-N., Dittmar, T., Gaupp, R. & Gleixner, G. The molecular composition of dissolved organic matter in forest soils as a function of pH and temperature. *PLoS One* **10**, e0119188 (2015).
48. Schmidt, M. W. I. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
49. Tao, F. *et al.* Microbial carbon use efficiency promotes global soil carbon storage. *Nature* **618**, 981–985 (2023).
50. Hedges, J. I. & Oades, J. M. Comparative organic geochemistries of soils and marine sediments. *Org. Geochem.* **27**, 319–361 (1997).
51. McCallister, S. L. & del Giorgio, P. A. Evidence for the respiration of ancient terrestrial organic C in northern temperate lakes and streams. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16963–16968 (2012).

52. Ward, N. D. *et al.* Degradation of terrestrially derived macromolecules in the Amazon River. *Nat. Geosci.* **6**, 530–533 (2013).
53. Ward, N. D. *et al. Integrative Research on Organic Matter Cycling Across Aquatic Gradients, 2nd Edition*. (Frontiers Media SA, 2020).
54. Bianchi, T. S. The role of terrestrially derived organic carbon in the coastal ocean: a changing paradigm and the priming effect. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19473–19481 (2011).
55. Spencer, R. G. M. *et al.* Photochemical degradation of dissolved organic matter and dissolved lignin phenols from the Congo River. *J. Geophys. Res.* **114**, (2009).
56. Moran, M. A. & Zepp, R. G. Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Limnol. Oceanogr.* **42**, 1307–1316 (1997).
57. Moran, M. A. & Zepp, R. G. UV radiation effects on microbes and microbial processes. *Microbial ecology of the*.
58. Blagodatsky, S., Blagodatskaya, E., Yuyukina, T. & Kuzyakov, Y. Model of apparent and real priming effects: Linking microbial activity with soil organic matter decomposition. *Soil Biol. Biochem.* **42**, 1275–1283 (2010).
59. Kuzyakov, Y. Priming effects: Interactions between living and dead organic matter. *Soil Biol. Biochem.* **42**, 1363–1371 (2010).
60. Lalonde, K., Mucci, A., Ouellet, A. & Gélinas, Y. Preservation of organic matter in sediments promoted by iron. *Nature* **483**, 198–200 (2012).
61. Arrieta, J. M. *et al.* Ocean chemistry. Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* **348**, 331–333 (2015).
62. Dittmar, T. *et al.* Enigmatic persistence of dissolved organic matter in the ocean. *Nature Reviews Earth & Environment* **2**, 570–583 (2021).
63. Berggren, M. *et al.* Unified understanding of intrinsic and extrinsic controls of dissolved organic carbon reactivity in aquatic ecosystems. *Ecology* **103**, e3763 (2022).
64. Kellerman, A. M., Kothawala, D. N., Dittmar, T. & Tranvik, L. J. Persistence of dissolved organic matter in lakes related to its molecular characteristics. *Nat. Geosci.* **8**, 454–457 (2015).

65. Kaplan, L. A. & Cory, R. M. Chapter 6 - Dissolved Organic Matter in Stream Ecosystems: Forms, Functions, and Fluxes of Watershed Tea. in *Stream Ecosystems in a Changing Environment* (eds. Jones, J. B. & Stanley, E. H.) 241–320 (Academic Press, 2016).
66. Hruška, J., Laudon, H. & Johnson, C. E. Acid/base character of organic acids in a boreal stream during snowmelt. *Water Resour.* doi:10.1029/2000WR900290.
67. Bishop, K., Seibert, J., Köhler, S. & Laudon, H. Resolving the Double Paradox of rapidly mobilized old water with highly variable responses in runoff chemistry. *Hydrol. Process.* **18**, 185–189 (2004).
68. Perdrial, J. N. *et al.* Stream water carbon controls in seasonally snow-covered mountain catchments: impact of inter-annual variability of water fluxes, catchment aspect and seasonal processes. *Biogeochemistry* vol. 118 273–290 Preprint at https://doi.org/10.1007/s10533-013-9929-y (2014).
69. Wetzel, R. G. Gradient-dominated ecosystems: sources and regulatory functions of dissolved organic matter in freshwater ecosystems. in *Dissolved Organic Matter in Lacustrine Ecosystems: Energy Source and System Regulator* (eds. Salonen, K., Kairesalo, T. & Jones, R. I.) 181–198 (Springer Netherlands, 1992).
70. Hornberger, G. M., Bencala, K. E. & McKnight, D. M. Hydrological controls on dissolved organic carbon during snowmelt in the Snake River near Montezuma, Colorado. *Biogeochemistry* **25**, 147–165 (1994).
71. Boyer, E. W., Hornberger, G. M. & Bencala, K. E. Response characteristics of DOC flushing in an alpine catchment. *Hydrological* doi:10.1002/(SICI)1099-1085(19971015)11:12<1635::AID-HYP494>3.0.CO;2-H.
72. Brooks, P. D., McKnight, D. M. & Bencala, K. E. Relationship between over-winter soil heterotrophic activity and DOC export in high elevation catchments. *Water Resour. Res.*
73. Hood, E., Williams, M. W. & McKnight, D. M. Sources of dissolved organic matter (DOM) in a Rocky Mountain stream using chemical fractionation and stable isotopes. *Biogeochemistry* **74**, 231–255 (2005).
74. Osburn, C. L. & Bianchi, T. S. *Linking Optical and Chemical Properties of Dissolved Organic Matter in Natural Waters*. (Frontiers Media SA, 2017).

75. Del Vecchio, R. & Blough, N. V. On the origin of the optical properties of humic substances. *Environ. Sci. Technol.* **38**, 3885–3891 (2004).
76. Burns, M. A., Barnard, H. R., Gabor, R. S., McKnight, D. M. & Brooks, P. D. Dissolved organic matter transport reflects hillslope to stream connectivity during snowmelt in a montane catchment. *Water Resour. Res.* **52**, 4905–4923 (2016).
77. Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).
78. Attermeyer, K., Premke, K., Hornick, T., Hilt, S. & Grossart, H.-P. Ecosystem-level studies of terrestrial carbon reveal contrasting bacterial metabolism in different aquatic habitats. *Ecology* **94**, 2754–2766 (2013).
79. Battin, T. J., Besemer, K., Bengtsson, M. M., Romani, A. M. & Packmann, A. I. The ecology and biogeochemistry of stream biofilms. *Nat. Rev. Microbiol.* **14**, 251–263 (2016).
80. Tranvik, L. J. *et al.* Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **54**, 2298–2314 (2009).
81. Kandasamy, S. & Nagender Nath, B. Perspectives on the Terrestrial Organic Matter Transport and Burial along the Land-Deep Sea Continuum: Caveats in Our Understanding of Biogeochemical Processes and Future Needs. *Frontiers in Marine Science* **3**, (2016).
82. Tanentzap, A. J. *et al.* Forests fuel fish growth in freshwater deltas. *Nat. Commun.* **5**, 4077 (2014).
83. Karlsson, J. *et al.* Terrestrial organic matter support of lake food webs: Evidence from lake metabolism and stable hydrogen isotopes of consumers. *Limnol. Oceanogr.* **57**, 1042–1048 (2012).
84. Jansson, M., Persson, L., De Roos, A. M., Jones, R. I. & Tranvik, L. J. Terrestrial carbon and intraspecific size-variation shape lake ecosystems. *Trends Ecol. Evol.* **22**, 316–322 (2007).
85. Creed, I. F. *et al.* Global change-driven effects on dissolved organic matter composition: Implications for food webs of northern lakes. *Glob. Chang. Biol.* **24**, 3692–3714 (2018).
86. Woodman, S. G. *et al.* Forest defoliator outbreaks alter nutrient cycling in northern waters. *Nat. Commun.* **12**, 6355 (2021).

87. Comte, J. & Del Giorgio, P. A. Composition influences the pathway but not the outcome of the metabolic response of bacterioplankton to resource shifts. *PLoS One* **6**, e25266 (2011).
88. Vannote, R. L., Minshall, G. W. & Cummins, K. W. The river continuum concept. *fisheries and aquatic …* (1980).
89. Creed, I. F., McKnight, D. M., Pellerin, B. A., Green, M. B. & Stackpoole, S. M. The river as a chemostat: Fresh perspectives on dissolved organic matter flowing down the river continuum. *Can. J. Fish. Aquat. Sci.* 150416143426006 (2015).
90. Worrall, F. *et al.* Trends in Dissolved Organic Carbon in UK Rivers and Lakes. *Biogeochemistry* **70**, 369–402 (2004).
91. Helms, J. R. *et al.* Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnol. Oceanogr.* **53**, 955–969 (2008).
92. Baker, A. Spectrophotometric discrimination of river dissolved organic matter. *Hydrol. Process.* **16**, 3203–3213 (2002).
93. Yamashita, Y. & Jaffé, R. Characterizing the interactions between trace metals and dissolved organic matter using excitation-emission matrix and parallel factor analysis. *Environ. Sci. Technol.* **42**, 7374–7379 (2008).
94. Huber, S. A., Balz, A., Abert, M. & Pronk, W. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography--organic carbon detection--organic nitrogen detection (LC-OCD-OND). *Water Res.* **45**, 879–885 (2011).
95. Abe, T., Maie, N. & Watanabe, A. Investigation of humic acid N with X-ray photoelectron spectroscopy: Effect of acid hydrolysis and comparison with 15N cross polarization/magic angle spinning nuclear magnetic resonance spectroscopy. *Organic Geochemistry* vol. 36 1490–1497 Preprint at https://doi.org/10.1016/j.orggeochem.2005.07.005 (2005).
96. Hernández, F., Sancho, J. V. & Pozo, O. J. Critical review of the application of liquid chromatography/mass spectrometry to the determination of pesticide residues in biological samples. *Anal. Bioanal. Chem.* **382**, 934–946 (2005).

97. Zark, M., Christoffers, J. & Dittmar, T. Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar. Chem.* **191**, 9–15 (2017).
98. Riedel, T. & Dittmar, T. A method detection limit for the analysis of natural organic matter via Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **86**, 8376–8382 (2014).
99. Schijf, J. & Zoll, A. M. When dissolved is not truly dissolved—The importance of colloids in studies of metal sorption on organic matter. *J. Colloid Interface Sci.* **361**, 137–147 (2011).
100. Wagner, S., Ding, Y. & Jaffé, R. A new perspective on the apparent solubility of dissolved black carbon. *Front. Earth Sci.* **5**, (2017).
101. Kim, S., Simpson, A. J., Kujawinski, E. B., Freitas, M. A. & Hatcher, P. G. High resolution electrospray ionization mass spectrometry and 2D solution NMR for the analysis of DOM extracted by C18 solid phase disk. *Org. Geochem.* **34**, 1325–1335 (2003).
102. Piccolo, A. THE SUPRAMOLECULAR STRUCTURE OF HUMIC SUBSTANCES. *Soil Science* vol. 166 810–832 Preprint at https://doi.org/10.1097/00010694-200111000-00007 (2001).
103. Hertkorn, N. *et al.* Characterization of a major refractory component of marine dissolved organic matter. *Geochim. Cosmochim. Acta* **70**, 2990–3010 (2006).
104. Qi, Y. *et al.* Deciphering dissolved organic matter by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS): from bulk to fractions and individuals. *Carbon Research* **1**, 3 (2022).
105. Hawkes, J. A., Patriarca, C., Sjöberg, P. J. R., Tranvik, L. J. & Bergquist, J. Extreme isomeric complexity of dissolved organic matter found across aquatic environments. *Limnol. Oceanogr. Lett.* **3**, 21–30 (2018).
106. Minor, E. C., Swenson, M. M., Mattson, B. M. & Oyler, A. R. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. *Environ. Sci. Process. Impacts* **16**, 2064–2079 (2014).
107. Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods* vol. 6 230–235 Preprint at https://doi.org/10.4319/lom.2008.6.230 (2008).

108. Lozano, D. C. P. *et al.* Pushing the analytical limits: new insights into complex mixtures using mass spectra segments of constant ultrahigh resolving power. *Chem. Sci.* **10**, 6966–6978 (2019).
109. Simon, C., Roth, V.-N., Dittmar, T. & Gleixner, G. Molecular Signals of Heterogeneous Terrestrial Environments Identified in Dissolved Organic Matter: A Comparative Analysis of Orbitrap and Ion Cyclotron Resonance Mass Spectrometers. *Front Earth Sci. Chin.* **6**, (2018).
110. Tanentzap, A. J. *et al.* Chemical and microbial diversity covary in fresh water to influence ecosystem functioning. *Proceedings of the National Academy of Sciences* vol. 116 24689–24695 Preprint at https://doi.org/10.1073/pnas.1904896116 (2019).
111. Mentges, A., Feenders, C., Seibt, M., Blasius, B. & Dittmar, T. Functional Molecular Diversity of Marine Dissolved Organic Matter Is Reduced during Degradation. *Frontiers in Marine Science* vol. 4 Preprint at https://doi.org/10.3389/fmars.2017.00194 (2017).
112. Noriega-Ortega, B. E. *et al.* Does the Chemodiversity of Bacterial Exometabolomes Sustain the Chemodiversity of Marine Dissolved Organic Matter? *Frontiers in Microbiology* vol. 10 Preprint at https://doi.org/10.3389/fmicb.2019.00215 (2019).
113. Danczak, R. E. *et al.* Using metacommunity ecology to understand environmental metabolomes. *Nat. Commun.* **11**, 6369 (2020).
114. Von Bertalanffy, L. The History and Status of General Systems Theory. *AMJ* **15**, 407–426 (1972).
115. Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423 (1948).
116. Hu, A., Meng, F., Tanentzap, A. J., Jang, K.-S. & Wang, J. Dark Matter Enhances Interactions within Both Microbes and Dissolved Organic Matter under Global Change. *Environ. Sci. Technol.* (2022) doi:10.1021/acs.est.2c05052.
117. Hu, A. *et al.* Ecological networks of dissolved organic matter and microorganisms under global change. *Nat. Commun.* **13**, 3600 (2022).

118. Mitschke, N., Vemulapalli, S. P. B. & Dittmar, T. NMR spectroscopy of dissolved organic matter: a review. *Environ. Chem. Lett.* (2022) doi:10.1007/s10311-022-01528-4.
119. Novak, J. M., Bertsch, P. M. & Mills, G. L. Carbon-13 nuclear magnetic resonance spectra of soil water-soluble organic carbon. *J. Environ. Qual.* **21**, 537–539 (1992).
120. Koch, B. P. & Dittmar, T. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* **20**, 926–932 (2006).
121. Merel, S. Critical assessment of the Kendrick mass defect analysis as an innovative approach to process high resolution mass spectrometry data for environmental applications. *Chemosphere* 137443 (2022).
122. Rivas-Ubach, A. *et al.* Moving beyond the van Krevelen Diagram: A New Stoichiometric Approach for Compound Classification in Organisms. *Anal. Chem.* **90**, 6152–6160 (2018).
123. Kim, S., Kramer, R. W. & Hatcher, P. G. Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Anal. Chem.* **75**, 5336–5344 (2003).
124. D'Andrilli, J., Cooper, W. T., Foreman, C. M. & Marshall, A. G. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. *Rapid Commun. Mass Spectrom.* **29**, 2385–2401 (2015).
125. Mann, B. F. *et al.* Indexing Permafrost Soil Organic Matter Degradation Using High-Resolution Mass Spectrometry. *PLoS One* **10**, e0130557 (2015).
126. Kothawala, D. N., Kellerman, A. M., Catalán, N. & Tranvik, L. J. Organic Matter Degradation across Ecosystem Boundaries: The Need for a Unified Conceptualization. *Trends Ecol. Evol.* **36**, 113–122 (2021).
127. Magurran, A. E. Q&A: What is biodiversity? *BMC Biol.* **8**, 145 (2010).
128. McDonough, L. K. *et al.* A new conceptual framework for the transformation of groundwater dissolved organic matter. *Nat. Commun.* **13**, 2153 (2022).
129. Hanski, I. & Professor in the Department of Ecology and Systematics Ilkka Hanski. *Metapopulation Ecology*. (OUP Oxford, 1999).
130. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. (Princeton University Press, 2011).

131. Food & Organization, A. FAOSTAT—Forestry database. (2013).
132. Campeau, A. *et al.* Multiple sources and sinks of dissolved inorganic carbon across Swedish streams, refocusing the lens of stable C isotopes. *Sci. Rep.* **7**, 9158 (2017).
133. Danneyrolles, V. *et al.* Stronger influence of anthropogenic disturbance than climate change on century-scale compositional changes in northern forests. *Nat. Commun.* **10**, 1265 (2019).
134. Kreutzweiser, D. P., Hazlett, P. W. & Gunn, J. M. Logging impacts on the biogeochemistry of boreal forest soils and nutrient export to aquatic systems: A review. *Environ. Rev.* **16**, 157–179 (2008).
135. Webster, K. L. *et al.* Long-term stream chemistry response to harvesting in a northern hardwood forest watershed experiencing environmental change. *For. Ecol. Manage.* **519**, 120345 (2022).
136. Laudon, H. *et al.* Response of Dissolved Organic Carbon following Forest Harvesting in a Boreal Forest. *Ambio* **38**, 381–386 (2009).
137. Kortelainen, P. & Saukkonen, S. Leaching of Nutrients, Organic Carbon and Iron from Finnish Forestry Land. *Water Air Soil Pollut.* **105**, 239–250 (1998).
138. Lamontagne, S., Carignan, R., D'Arcy, P., Prairie, Y. T. & Paré, D. Element export in runoff from eastern Canadian Boreal Shield drainage basins following forest harvesting and wildfires. *Can. J. Fish. Aquat. Sci.* **57**, 118–128 (2000).
139. Schelker, J. Forestry impact on water quality: a landscape perspective on dissolved organic carbon. (2013).
140. Food and Agriculture Organization of the United Nations. *Global Soil Organic Carbon Map – GSOCmap v.1.6: Technical report*. (Food & Agriculture Org., 2022).
141. Batjes, N. H. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **269**, 61–68 (2016).
142. Buttle, J. M. *et al.* Hydrologic response to and recovery from differing silvicultural systems in a deciduous forest landscape with seasonal snow cover. *Journal of Hydrology* vol. 557 805–825 Preprint at https://doi.org/10.1016/j.jhydrol.2018.01.006 (2018).

143. Jeffries, D. S. & Foster, N. W. The Turkey Lakes Watershed Study: Milestones and Prospects. *Ecosystems* **4**, 501–502 (2001).
144. Eklöf, K., Lidskog, R. & Bishop, K. Managing Swedish forestry's impact on mercury in fish: Defining the impact and mitigation measures. *Ambio* vol. 45 163–174 Preprint at https://doi.org/10.1007/s13280-015-0752-7 (2016).
145. Feller, M. C. FOREST HARVESTING AND STREAMWATER INORGANIC CHEMISTRY IN WESTERN NORTH AMERICA: A REVIEW. *Journal of the American Water Resources Association* vol. 41 785–811 Preprint at https://doi.org/10.1111/j.1752-1688.2005.tb03771.x (2005).
146. Ågren, A. M., Lidberg, W. & Ring, E. Mapping Temporal Dynamics in a Forest Stream Network—Implications for Riparian Forest Management. *For. Trees Livelihoods* **6**, 2982–3001 (2015).
147. Laudon, H. *et al.* The role of biogeochemical hotspots, landscape heterogeneity, and hydrological connectivity for minimizing forestry effects on water quality. *Ambio* **45 Suppl 2**, 152–162 (2016).
148. Moore, R., Spittlehouse, D. L. & Story, A. RIPARIAN MICROCLIMATE AND STREAM TEMPERATURE RESPONSE TO FOREST HARVESTING: A REVIEW. *Journal of the American Water Resources Association* vol. 41 813–834 Preprint at https://doi.org/10.1111/j.1752-1688.2005.tb04465.x (2005).
149. Troendle, C. A. & King, R. M. The effect of timber harvest on the Fool Creek watershed, 30 years later. *Water Resour. Res.* doi:10.1029/WR021i012p01915.
150. Collins, B. J., Rhoades, C. C., Hubbard, R. M. & Battaglia, M. A. Tree regeneration and future stand development after bark beetle infestation and harvesting in Colorado lodgepole pine stands. *For. Ecol. Manage.* **261**, 2168–2175 (2011).
151. Lotan, J. E., Chudnoff, M. & Perry, D. A. *Ecology and Regeneration of Lodgepole Pine*. (U.S. Department of Agriculture, Forest Service, 1983).
152. Rhoades, C. C., Hubbard, R. M. & Elder, K. A Decade of Streamwater Nitrogen and Forest Dynamics after a Mountain Pine Beetle Outbreak at the Fraser Experimental Forest, Colorado. *Ecosystems* vol. 20 380–392 Preprint at https://doi.org/10.1007/s10021-016-0027-6 (2017).

153. Chatterjee, A., Vance, G. F., Pendall, E. & Stahl, P. D. Timber harvesting alters soil carbon mineralization and microbial community structure in coniferous forests. *Soil Biol. Biochem.* **40**, 1901–1907 (2008).
154. Douglas, R. B., Parker, V. T. & Cullings, K. W. Belowground ectomycorrhizal community structure of mature lodgepole pine and mixed conifer stands in Yellowstone National Park. *For. Ecol. Manage.* **208**, 303–317 (2005).
155. Evans, L. R., Pierson, D. & Lajtha, K. Dissolved organic carbon production and flux under long-term litter manipulations in a Pacific Northwest old-growth forest. *Biogeochemistry* vol. 149 75–86 Preprint at https://doi.org/10.1007/s10533-020-00667-6 (2020).
156. Fakhraei, H., Fahey, T. J. & Driscoll, C. T. The Biogeochemical Response of Nitrate and Potassium to Landscape Disturbance in Watersheds of the Hubbard Brook Experimental Forest, New Hampshire, USA. *Forest-Water Interactions* 537–563 Preprint at https://doi.org/10.1007/978-3-030-26086-6_22 (2020).
157. Drake, T. W. *et al.* Mobilization of aged and biolabile soil carbon by tropical deforestation. *Nat. Geosci.* **12**, 541–546 (2019).
158. Kellerman, A. M. *et al.* Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. *Environ. Sci. Technol.* **52**, 2538–2548 (2018).
159. Roth, V.-N., Dittmar, T., Gaupp, R. & Gleixner, G. Latitude and pH driven trends in the molecular composition of DOM across a north south transect along the Yenisei River. *Geochim. Cosmochim. Acta* **123**, 93–105 (2013).
160. Roth, V.-N., Dittmar, T., Gaupp, R. & Gleixner, G. Ecosystem-specific composition of dissolved organic matter. *Vadose Zone J.* **13**, vzj2013.09.0162 (2014).
161. Seifert, A.-G. *et al.* Comparing molecular composition of dissolved organic matter in soil and stream water: Influence of land use and chemical characteristics. *Sci. Total Environ.* **571**, 142–152 (2016).
162. Hawkes, J. A. *et al.* Regional diversity of complex dissolved organic matter across forested hemiboreal headwater streams. *Sci. Rep.* **8**, 1–11 (2018).
163. Kothawala, D. N. *et al.* 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 (2015).

164. Friedlingstein, P. *et al.* Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
165. *IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.
166. Nakhavali, M. *et al.* Leaching of dissolved organic carbon from mineral soils plays a significant role in the terrestrial carbon balance. *Glob. Chang. Biol.* (2020) doi:10.1111/gcb.15460.
167. Meyers-Schulte, K. J. & Hedges, J. I. Molecular evidence for a terrestrial component of organic matter dissolved in ocean water. *Nature* **321**, 61–63 (1986).
168. Kellerman, A. M., Dittmar, T., Kothawala, D. N. & Tranvik, L. J. Chemodiversity of dissolved organic matter in lakes driven by climate and hydrology. *Nat. Commun.* **5**, 3804 (2014).
169. Mosher, J. J., Kaplan, L. A., Podgorski, D. C., McKenna, A. M. & Marshall, A. G. Longitudinal shifts in dissolved organic matter chemogeography and chemodiversity within headwater streams: a river continuum reprise. *Biogeochemistry* **124**, 371–385 (2015).
170. Seidel, M., Vemulapalli, S. P. B., Mathieu, D. & Dittmar, T. Marine Dissolved Organic Matter Shares Thousands of Molecular Formulae Yet Differs Structurally across Major Water Masses. *Environ. Sci. Technol.* **56**, 3758–3769 (2022).
171. *Applying the core-satellite species concept: Characteristics of rare and common riverine dissolved organic matter*.
172. Middelburg, J. J. A simple rate model for organic matter decomposition in marine sediments. *Geochim. Cosmochim. Acta* **53**, 1577–1581 (1989).
173. Boudreau, B. P., Arnosti, C., Jørgensen, B. B. & Canfield, D. E. Comment on “Physical model for the decay and preservation of marine organic carbon.” *Science* vol. 319 1616; author reply 1616 (2008).
174. Del Giorgio, P. A., Davis, J., Findlay, S. E. G. & Sinsabaugh, R. L. Aquatic ecosystems: interactivity of dissolved organic matter. *Elsevier Science (USA)* 400–424 Preprint at (2003).

175. Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T. & Kattner, G. Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Geochim. Cosmochim. Acta* **69**, 3299–3308 (2005).
176. Lam, B. *et al.* Major Structural Components in Freshwater Dissolved Organic Matter. *Environmental Science & Technology* vol. 41 8240–8247 Preprint at https://doi.org/10.1021/es0713072 (2007).
177. Lechtenfeld, O. J. *et al.* Molecular transformation and degradation of refractory dissolved organic matter in the Atlantic and Southern Ocean. *Geochim. Cosmochim. Acta* **126**, 321–337 (2014).
178. Roth, V.-N. *et al.* Persistence of dissolved organic matter explained by molecular changes during its passage through soil. *Nat. Geosci.* **12**, 755–761 (2019).
179. Hugouvieux-Cotte-Pattat, N., Condemine, G. & Shevchik, V. E. Bacterial pectate lyases, structural and functional diversity. *Environ. Microbiol. Rep.* **6**, 427–440 (2014).
180. Bradford, M. A., Keiser, A. D., Davies, C. A., Mersmann, C. A. & Strickland, M. S. Empirical evidence that soil carbon formation from plant inputs is positively related to microbial growth. *Biogeochemistry* **113**, 271–281 (2013).
181. Vaishnavi, J. & Osborne, W. J. Chapter 3 - Microbial volatiles: small molecules with an important role in intra- and interbacterial genus interactions-quorum sensing. in *Volatiles and Metabolites of Microbes* (eds. Kumar, A., Singh, J. & Samuel, J.) 35–50 (Academic Press, 2021).
182. Ladygina, N., Dedyukhina, E. G. & Vainshtein, M. B. A review on microbial synthesis of hydrocarbons. *Process Biochem.* **41**, 1001–1014 (2006).
183. Dungait, J. A. J., Hopkins, D. W., Gregory, A. S. & Whitmore, A. P. Soil organic matter turnover is governed by accessibility not recalcitrance. *Glob. Chang. Biol.* **18**, 1781–1796 (2012).
184. Lynch, L. M. *et al.* River channel connectivity shifts metabolite composition and dissolved organic matter chemistry. *Nat. Commun.* **10**, 459 (2019).
185. Fontaine, S. *et al.* Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* **450**, 277–280 (2007).

186. Zakem, E. J., Cael, B. B. & Levine, N. M. A unified theory for organic matter accumulation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
187. Xenopoulos, M. A., Downing, J. A., Kumar, M. D., Menden-Deuer, S. & Voss, M. Headwaters to oceans: Ecological and biogeochemical contrasts across the aquatic continuum. *Limnol. Oceanogr.* **62**, S3–S14 (2017).
188. Merder, J. *et al.* Improved Mass Accuracy and Isotope Confirmation through Alignment of Ultrahigh-Resolution Mass Spectra of Complex Natural Mixtures. *Anal. Chem.* **92**, 2558–2565 (2020).
189. Huber, S. A., Balz, A., Abert, M. & Pronk, W. Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography – organic carbon detection – organic nitrogen detection (LC-OCD-OND). *Water Research* vol. 45 879–885 Preprint at https://doi.org/10.1016/j.watres.2010.09.023 (2011).
190. Hawkes, J. A., Hansen, C. T., Goldhammer, T., Bach, W. & Dittmar, T. Molecular alteration of marine dissolved organic matter under experimental hydrothermal conditions. *Geochim. Cosmochim. Acta* **175**, 68–85 (2016).
191. Post-Beittenmiller, D. BIOCHEMISTRY AND MOLECULAR BIOLOGY OF WAX PRODUCTION IN PLANTS. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 405–430 (1996).
192. Haslam, E. *Plant Polyphenols: Vegetable Tannins Revisited*. (CUP Archive, 1989).
193. Taipale, S. J. *et al.* Terrestrial carbohydrates support freshwater zooplankton during phytoplankton deficiency. *Sci. Rep.* **6**, 30897 (2016).
194. Langeveld, J. *et al.* Estimating dissolved carbon concentrations in global soils: a global database and model. *SN Applied Sciences* **2**, 1626 (2020).
195. Neff, J. C. & Asner, G. P. Dissolved Organic Carbon in Terrestrial Ecosystems: Synthesis and a Model. *Ecosystems* **4**, 29–48 (2001).
196. Rumpel, C. & Kögel-Knabner, I. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**, 143–158 (2011).
197. Yu, W. H. *et al.* Adsorption of proteins and nucleic acids on clay minerals and their interactions: A review. *Appl. Clay Sci.* **80–81**, 443–452 (2013).
198. Huggett, R. J. *Fundamentals of Geomorphology*. (Routledge, 2007).

199. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**, 41 (2013).
200. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).
201. Lombard, V. *et al.* A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem. J.* **432**, 437–444 (2010).
202. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J* **382**, 769–781 (2004).
203. Shen, Y., Chapelle, F. H., Strom, E. W. & Benner, R. Origins and bioavailability of dissolved organic matter in groundwater. *Biogeochemistry* **122**, 61–78 (2015).
204. Jaffé, R. *et al.* Global charcoal mobilization from soils via dissolution and riverine transport to the oceans. *Science* **340**, 345–347 (2013).
205. Haegeman, B. *et al.* Robust estimation of microbial diversity in theory and in practice. *ISME J.* **7**, 1092–1101 (2013).
206. Jouanin, L. & Lapierre, C. *Lignins: Biosynthesis, Biodegradation and Bioengineering*. (Academic Press, 2012).
207. Saidy, A. R., Smernik, R. J., Baldock, J. A., Kaiser, K. & Sanderman, J. The sorption of organic carbon onto differing clay minerals in the presence and absence of hydrous iron oxide. *Geoderma* **209–210**, 15–21 (2013).
208. Conacher, A. J. & Dalrymple, J. B. The nine unit landsurface model: an approach to pedogeomorphic research. *Geoderma* (1977).
209. Nobre, A. D. *et al.* Height Above the Nearest Drainage – a hydrologically relevant new terrain model. *J. Hydrol.* **404**, 13–29 (2011).
210. Nobre, A. D. *et al.* HAND contour: a new proxy predictor of inundation extent. *Hydrol. Process.* **30**, 320–333 (2016).
211. Seeberg-Elverfeldt, J., Schlüter, M., Feseker, T. & Kölling, M. Rhizon sampling of porewaters near the sediment-water interface of aquatic systems. *Limnol. Oceanogr. Methods* **3**, 361–371 (2005).
212. Merder, J. *et al.* ICBM-OCEAN: Processing Ultrahigh-Resolution Mass Spectrometry Data of Complex Molecular Mixtures. *Anal. Chem.* **92**, 6832–6838 (2020).

213. Bååth, E., Pettersson, M. & Söderberg, K. H. Adaptation of a rapid and economical microcentrifugation method to measure thymidine and leucine incorporation by soil bacteria. *Soil Biol. Biochem.* **33**, 1571–1574 (2001).
214. Bååth, E. Thymidine and leucine incorporation in soil bacteria with different cell size. *Microb. Ecol.* **27**, 267–278 (1994).
215. Bade, D., Houser, J. & Scanga, S. Methods of the cascading trophic interactions project. *Center for Limnology, University of Wisconsin--Madison* (1998).
216. Findlay, S. E. G. & Parr, T. B. Chapter 24 - Dissolved Organic Matter. in *Methods in Stream Ecology (Third Edition)* (eds. Lamberti, G. A. & Hauer, F. R.) 21–36 (Academic Press, 2017).
217. Saiya-Cork, K. R., Sinsabaugh, R. L. & Zak, D. R. The effects of long term nitrogen deposition on extracellular enzyme activity in an Acer saccharum forest soil. *Soil Biol. Biochem.* **34**, 1309–1315 (2002).
218. Khalili, B., Weihe, C., Kimball, S., Schmidt, K. T. & Martiny, J. B. H. Optimization of a Method To Quantify Soil Bacterial Abundance by Flow Cytometry. *mSphere* **4**, (2019).
219. Frossard, A., Hammes, F. & Gessner, M. O. Flow Cytometric Assessment of Bacterial Abundance in Soils, Sediments and Sludge. *Front. Microbiol.* **7**, 903 (2016).
220. Rheault, K. *et al.* Plant Genotype Influences Physicochemical Properties of Substrate as Well as Bacterial and Fungal Assemblages in the Rhizosphere of Balsam Poplar. *Front. Microbiol.* **11**, 575625 (2020).
221. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* vol. 34 i884–i890 Preprint at https://doi.org/10.1093/bioinformatics/bty560 (2018).
222. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Research* vol. 27 824–834 Preprint at https://doi.org/10.1101/gr.213959.116 (2017).
223. Arkin, A. P. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
224. Nikolenko, S. I., Korobeynikov, A. I. & Alekseyev, M. A. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14 Suppl 1**, S7 (2013).

225. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
226. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
227. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
228. Love, M., Anders, S. & Huber, W. Differential analysis of count data--the DESeq2 package. *Genome Biol.* **15**, 10–1186 (2014).
229. Lenth, R. emmeans: estimated marginal means, aka least-squares means. R package version 1.4. 7. 2020. Preprint at (2022).
230. Rao, C. R. The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* **26**, 329–358 (1964).
231. Peres-Neto, P. R., Legendre, P., Dray, S. & Borcard, D. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**, 2614–2625 (2006).
232. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.* **196**, 483–493 (2006).
233. Blanchet, F. G., Legendre, P. & Borcard, D. Forward selection of explanatory variables. *Ecology* **89**, 2623–2632 (2008).
234. Lai, J., Zou, Y., Zhang, J. & Peres-Neto, P. R. Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. *Methods Ecol. Evol.* (2022) doi:10.1111/2041-210x.13800.
235. Nepal, P., Johnston, C. M. T. & Ganguly, I. Effects on Global Forests and Wood Product Markets of Increased Demand for Mass Timber. *Sustain. Sci. Pract. Policy* **13**, 13943 (2021).
236. Searchinger, T., James, O., Dumas, P., Kastner, T. & Wirsenius, S. EU climate plan sacrifices carbon storage and biodiversity for bioenergy. *Nature* **612**, 27–30 (2022).
237. Mishra, A. *et al.* Land use change and carbon emissions of a transformation to timber cities. *Nat. Commun.* **13**, 4889 (2022).

238. Meyer, J. L. & Tate, C. M. The effects of watershed disturbance on dissolved organic carbon dynamics of a stream. *Ecology* **64**, 33–44 (1983).
239. Cronan, C. S., Piampiano, J. T. & Patterson, H. H. Influence of land use and hydrology on exports of carbon and nitrogen in a Maine river basin. *J. Environ. Qual.* **28**, 953–961 (1999).
240. Evans, C. D. *et al.* Variability in organic carbon reactivity across lake residence time and trophic gradients. *Nat. Geosci.* **10**, 832–835 (2017).
241. Hu, A. *et al.* Microbial and Environmental Processes Shape the Link between Organic Matter Functional Traits and Composition. *Environ. Sci. Technol.* **56**, 10504–10516 (2022).
242. Freeman, E. C., Creed, I. F., Jones, B. & Bergström, A.-K. Global changes may be promoting a rise in select cyanobacteria in nutrient-poor northern lakes. *Glob. Chang. Biol.* **26**, 4966–4987 (2020).
243. Sheridan, E. A. *et al.* Plastic pollution fosters more microbial growth in lakes than natural organic matter. *Nat. Commun.* **13**, 4175 (2022).
244. Flerus, R. *et al.* A molecular perspective on the ageing of marine dissolved organic matter. *Biogeosciences* **9**, 1935–1955 (2012).
245. Kim, S., Kaplan, L. A. & Hatcher, P. G. Biodegradable dissolved organic matter in a temperate and a tropical stream determined from ultra-high resolution mass spectrometry. *Limnol. Oceanogr.* **51**, 1054–1063 (2006).
246. Textor, S. R., Wickland, K. P., Podgorski, D. C., Johnston, S. E. & Spencer, R. G. M. Dissolved Organic Carbon Turnover in Permafrost-Influenced Watersheds of Interior Alaska: Molecular Insights and the Priming Effect. *Front Earth Sci. Chin.* **7**, (2019).
247. Raymond, P. A. & Bauer, J. E. Riverine export of aged terrestrial organic matter to the North Atlantic Ocean. *Nature* **409**, 497–500 (2001).
248. Mayorga, E. *et al.* Young organic matter as a source of carbon dioxide outgassing from Amazonian rivers. *Nature* **436**, 538–541 (2005).
249. Barnes, R. T., Butman, D. E., Wilson, H. F. & Raymond, P. A. Riverine Export of Aged Carbon Driven by Flow Path Depth and Residence Time. *Environ. Sci. Technol.* **52**, 1028–1035 (2018).
250. Mann, P. J. *et al.* Utilization of ancient permafrost carbon in headwaters of Arctic fluvial networks. *Nat. Commun.* **6**, 7856 (2015).

251. Hood, E. *et al.* Glaciers as a source of ancient and labile organic matter to the marine environment. *Nature* **462**, 1044–1047 (2009).
252. Moore, S. *et al.* Deep instability of deforested tropical peatlands revealed by fluvial organic carbon fluxes. *Nature* **493**, 660–663 (2013).
253. Butman, D. E., Wilson, H. F., Barnes, R. T., Xenopoulos, M. A. & Raymond, P. A. Increased mobilization of aged carbon to rivers by human disturbance. *Nat. Geosci.* **8**, 112–116 (2014).
254. Lu, Y. H. *et al.* Effects of land use on sources and ages of inorganic and organic carbon in temperate headwater streams. *Biogeochemistry* **119**, 275–292 (2014).
255. Freeman, E. C. *et al.* Universal microbial reworking of dissolved organic matter along environmental gradients. *bioRxiv* 2022.11.23.517109 (2022) doi:10.1101/2022.11.23.517109.
256. Fasching, C., Ulseth, A. J., Schelker, J., Steniczka, G. & Battin, T. J. Hydrology controls dissolved organic matter export and composition in an Alpine stream and its hyporheic zone. *Limnol. Oceanogr.* **61**, 558–571 (2016).
257. Cole, J. J. *et al.* Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. *Ecosystems* **10**, 172–185 (2007).
258. Butman, D. & Raymond, P. A. Significant efflux of carbon dioxide from streams and rivers in the United States. *Nat. Geosci.* **4**, 839–842 (2011).
259. McLaughlin, J. W., Liu, G., Jurgensen, M. F. & Gale, M. R. Organic carbon characteristics in a spruce swamp five years after harvesting. *Soil Sci. Soc. Am. J.* **60**, 1228–1236 (1996).
260. Hyvönen, R., Olsson, B. A., Lundkvist, H. & Staaf, H. Decomposition and nutrient release from Picea abies (L.) Karst. and Pinus sylvestris L. logging residues. *For. Ecol. Manage.* **126**, 97–112 (2000).
261. Olchev, A., Radler, K., Sogachev, A., Panferov, O. & Gravenhorst, G. Application of a three-dimensional model for assessing effects of small clear-cuttings on radiation and soil temperature. *Ecol. Modell.* **220**, 3046–3056 (2009).
262. Schelker, J., Grabs, T., Bishop, K. & Laudon, H. Drivers of increased organic carbon concentrations in stream water following forest disturbance: Separating effects of changes in flow pathways and soil warming. *J. Geophys. Res. Biogeosci.* **118**, 1814–1827 (2013).

263. Sebestyen, S. D., Boyer, E. W. & Shanley, J. B. Sources, transformations, and hydrological processes that control stream nitrate and dissolved organic matter concentrations during snowmelt in an upland forest. *Water Resour.* (2008).
264. Seibert, Grabs, Köhler & Laudon. Linking soil-and stream-water chemistry based on a Riparian Flow-Concentration Integration Model. *Hydrol. Earth Syst. Sci.*
265. Gommet, C. *et al.* Spatiotemporal patterns and drivers of terrestrial dissolved organic carbon (DOC) leaching into the European river network. *Earth Syst. Dyn.* **13**, 393–418 (2022).
266. Ohno, T. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. *Environ. Sci. Technol.* **36**, 742–746 (2002).
267. Aiken, G. Fluorescence and Dissolved Organic Matter. in *Aquatic Organic Matter Fluorescence* 35–74 (Cambridge University Press, 2014).
268. Braga, L. P. P. *et al.* Viruses direct carbon cycling in lake sediments under global change. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202261119 (2022).
269. Thurman, E. M. Humic substances in groundwater. https://agris.fao.org/agris-search/search.do?recordID=US880370288.
270. Findlay, S. & Sinsabaugh, R. L. *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*. (Academic Press, 2003).
271. Benner, R., Fogel, M. L., Sprague, E. K. & Hodson, R. E. Depletion of 13C in lignin and its implications for stable carbon isotope studies. *Nature* **329**, 708–710 (1987).
272. Benner, R. & Kaiser, K. Biological and photochemical transformations of amino acids and lignin phenols in riverine dissolved organic matter. *Biogeochemistry* **102**, 209–222 (2011).
273. Moran, M. A. & Hodson, R. E. Dissolved humic substances of vascular plant origin in a coastal marine environment. *Limnol. Oceanogr.* **39**, 762–771 (1994).
274. Mostovaya, A., Hawkes, J. A., Dittmar, T. & Tranvik, L. J. Molecular Determinants of Dissolved Organic Matter Reactivity in Lake Water. *Front Earth Sci. Chin.* **5**, (2017).
275. Mostovaya, A., Koehler, B., Guillemette, F., Brunberg, A.-K. & Tranvik, L. J. Effects of compositional changes on reactivity continuum and decomposition

kinetics of lake dissolved organic matter. *J. Geophys. Res. Biogeosci.* **121**, 1733–1746 (2016).

276. Stubbins, A. *et al.* Illuminated darkness: Molecular signatures of Congo River dissolved organic matter and its photochemical alteration as revealed by ultrahigh precision mass spectrometry. *Limnol. Oceanogr.* **55**, 1467–1477 (2010).
277. Zhang, X. *et al.* Importance of lateral flux and its percolation depth on organic carbon export in Arctic tundra soil: Implications from a soil leaching experiment. *J. Geophys. Res. Biogeosci.* **122**, 796–810 (2017).
278. Hillel, D. Soil Dynamics. *Environmental Soil Physics* 341–382 Preprint at https://doi.org/10.1016/b978-012348525-0/50015-x (1998).
279. Horn, R., Vossbrink, J., Peth, S. & Becker, S. Impact of modern forest vehicles on soil physical properties. *For. Ecol. Manage.* **248**, 56–63 (2007).
280. Vossbrink, J. & Horn, R. Modern forestry vehicles and their impact on soil physical properties. *Eur. J. For. Res.* **123**, 259–267 (2004).
281. Williamson, J. R. & Neilsen, W. A. The influence of forest site on rate and extent of soil compaction and profile disturbance of skid trails during ground-based harvesting. *Can. J. For. Res.* **30**, 1196–1205 (2000).
282. Schoenholtz, S. H. HYDROLOGY | Impacts of Forest Management on Water Quality. *Encyclopedia of Forest Sciences* 377–387 Preprint at https://doi.org/10.1016/b0-12-145160-7/00209-x (2004).
283. Lehmann, J. *et al.* Persistence of soil organic carbon caused by functional complexity. *Nat. Geosci.* **13**, 529–534 (2020).
284. Mooney, H. A. & Cleland, E. E. The evolutionary impact of invasive species. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5446–5451 (2001).
285. Castañeda-Gómez, L. *et al.* Soil organic matter molecular composition with long-term detrital alterations is controlled by site-specific forest properties. *Glob. Chang. Biol.* **n/a**, (2022).
286. Lajtha, K. *et al.* The detrital input and removal treatment (DIRT) network: Insights into soil carbon stabilization. *Sci. Total Environ.* **640–641**, 1112–1120 (2018).
287. Emilson, E. J. S. *et al.* Climate-driven shifts in sediment chemistry enhance methane production in northern lakes. *Nat. Commun.* **9**, 1801 (2018).

288. Cristan, R. *et al.* Effectiveness of forestry best management practices in the United States: Literature review. *For. Ecol. Manage.* **360**, 133–151 (2016).
289. Beschta, R. L. & Jackson, W. L. Forest Practices and Sediment Production in the Alsea Watershed Study. in *Hydrological and Biological Responses to Forest Practices: The Alsea Watershed Study* (ed. Stednick, J. D.) 55–66 (Springer New York, 2008).
290. Brown, K. R., Michael Aust, W. & McGuire, K. J. Sediment delivery from bare and graveled forest road stream crossing approaches in the Virginia Piedmont. *For. Ecol. Manage.* **310**, 836–846 (2013).
291. Rowe, J. S. & Others. Forest regions of Canada. *Forest regions of Canada.* (1972).
292. Webster, K. L. *et al.* Turkey Lakes Watershed, Ontario, Canada: 40 years of interdisciplinary whole-ecosystem research. *Hydrol. Process.* **35**, (2021).
293. Foster, N. W., Beall, F. D. & Kreutzweiser, D. P. The role of forests in regulating water: The Turkey Lakes Watershed case study. *The Forestry Chronicle* vol. 81 142–148 Preprint at https://doi.org/10.5558/tfc81142-1 (2005).
294. Foster, N. *et al.* Heterogeneity in soil nitrogen within first-order forested catchments at the Turkey Lakes Watershed. *Can. J. For. Res.* **35**, 797–805 (2005).
295. Landry, G. *et al.* Mitigation Potential of Ecosystem-Based Forest Management under Climate Change: A Case Study in the Boreal-Temperate Forest Ecotone. *For. Trees Livelihoods* **12**, 1667 (2021).
296. Boisvert-Marsh, L., Périé, C. & de Blois, S. Shifting with climate? Evidence for recent changes in tree species distribution at high latitudes. *Ecosphere* **5**, art83 (2014).
297. Jeffries, D. S., Kelso, J. R. M. & Morrison, I. K. Physical, chemical, and biological characteristics of the Turkey Lakes Watershed, central Ontario, Canada. *Can. J. Fish. Aquat. Sci.* **45**, s3–s13 (1988).
298. Lindsay, J. B. The whitebox geospatial analysis tools project and open-access GIS. https://jblindsay.github.io/ghrg/pubs/LindsayGISRUK2014.pdf.
299. O'Callaghan, J. F. & Mark, D. M. The extraction of drainage networks from digital elevation data. *Comput. Vis. Graph. Image Process.* **27**, 247 (1984).

300. Pennock, D. J., Zebarth, B. J. & De Jong, E. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. *Geoderma* **40**, 297–315 (1987).
301. Rennó, C. D. *et al.* HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sens. Environ.* **112**, 3469–3481 (2008).
302. *stand-amp-site-guide.pdf*.
303. Pucher, M. PARAFAC analysis of EEM data to separate DOM components in R. *staRdom: spectroscopic analysis of dissolved organic matter in R. Retrieved from. https://cran. r-project. o r g/web/packages/staRdom/vignettes/PARAFAC_analysis_of_EEM. html# references. Accesed on March 23th* (2020).
304. Göttlein, A., Hell, U. & Blasek, R. A system for microscale tensiometry and lysimetry. *Geoderma* **69**, 147–156 (1996).
305. Vetter, W. F. W. McLafferty, F. Turecek. Interpretation of mass spectra. Fourth edition (1993). University Science Books, Mill Valley, California. *Biological Mass Spectrometry* vol. 23 379–379 Preprint at https://doi.org/10.1002/bms.1200230614 (1994).
306. LaRowe, D. E. & Van Cappellen, P. Degradation of natural organic matter: A thermodynamic analysis. *Geochim. Cosmochim. Acta* **75**, 2030–2042 (2011).
307. Koch, B. P., Dittmar, T., Witt, M. & Kattner, G. Fundamentals of molecular formula assignment to ultrahigh resolution mass data of natural organic matter. *Anal. Chem.* **79**, 1758–1763 (2007).
308. Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M. & Inouye, B. D. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* **2**, art24 (2011).
309. Leach, J. A., Hudson, D. T. & Moore, R. D. Assessing stream temperature response and recovery for different harvesting systems in northern hardwood forests using 40 years of spot measurements. *Hydrol. Process.* **36**, (2022).
310. Beall, F. D., Semkin, R. G. & Jeffries, D. S. Trends in the Output of First-order Basins at Turkey Lakes Watershed, 1982-96. *Ecosystems* vol. 4 514–526 Preprint at https://doi.org/10.1007/s10021-001-0025-0 (2001).

311. Morrison, I. K. & Hogan, G. D. Trace element distribution within the tree phytomass and forest floor of a tolerant hardwood stand, Algoma, Ontario. *Water Air Soil Pollut.* **31**, 493-500. (1986).
312. Lamlom, S. H. & Savidge, R. A. Carbon content variation in boles of mature sugar maple and giant sequoia. *Tree Physiol.* **26**, 459–468 (2006).
313. Wauchope, H. S. *et al.* Evaluating Impact Using Time-Series Data. *Trends Ecol. Evol.* **36**, 196–205 (2021).
314. Pinheiro, J. nlme : Linear and nonlinear mixed effects models. R package version 3.1-96. *http://cran.r-project.org/web/packages/nlme/* (2009).
315. Team, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. *http://www. R-project. org/* (2013).
316. Magnusson, A., Skaug, H., Nielsen, A. & Berg, C. glmmTMB: generalized linear mixed models using template model builder. *R package version 0.1*.
317. Janzen, H. H. Carbon cycling in earth systems—a soil science perspective. *Agric. Ecosyst. Environ.* **104**, 399–417 (2004).
318. Mayer, M. *et al.* Tamm Review: Influence of forest management activities on soil organic carbon stocks: A knowledge synthesis. *For. Ecol. Manage.* **466**, 118127 (2020).
319. Wei, X., Shao, M., Gale, W. & Li, L. Global pattern of soil carbon losses due to the conversion of forests to agricultural land. *Scientific Reports* vol. 4 Preprint at https://doi.org/10.1038/srep04062 (2015).
320. Mishra, A. *et al.* Estimating global land system impacts of timber plantations using MAgPIE 4.3.5. *Geosci. Model Dev.* **14**, 6467–6494 (2021).
321. Erika C. Freeman, Erik J.S Emilson, Andrew Tanentzap. Logging disrupts the ecology of molecules in headwater streams. *Nat. Geosci.*
322. Bro, R. PARAFAC. Tutorial and applications. *Chemometrics Intellig. Lab. Syst.* **38**, 149–171 (1997).
323. Stedmon, C. A. & Bro, R. Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnol. Oceanogr. Methods* **6**, 572–579 (2008).

324. Stedmon, C. A., Markager, S. & Bro, R. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Mar. Chem.* **82**, 239–254 (2003).
325. Cory, R. M. & McKnight, D. M. Fluorescence Spectroscopy Reveals Ubiquitous Presence of Oxidized and Reduced Quinones in Dissolved Organic Matter. *Environmental Science & Technology* vol. 39 8142–8149 Preprint at https://doi.org/10.1021/es0506962 (2005).
326. Gonsior, M., Schmitt-Kopplin, P. & Bastviken, D. Depth-dependent molecular composition and photo-reactivity of dissolved organic matter in a boreal lake under winter and summer conditions. *Biogeosciences* **10**, 6945–6956 (2013).
327. Gonsior, M. *et al.* Characterization of dissolved organic matter across the Subtropical Convergence off the South Island, New Zealand. *Mar. Chem.* **123**, 99–110 (2011).
328. Herzsprung, P. *et al.* Variations of DOM quality in inflows of a drinking water reservoir: linking of van Krevelen diagrams with EEMF spectra by rank correlation. *Environ. Sci. Technol.* **46**, 5511–5518 (2012).
329. Stubbins, A. *et al.* What's in an EEM? Molecular signatures associated with dissolved organic fluorescence in boreal Canada. *Environ. Sci. Technol.* **48**, 10598–10606 (2014).
330. Wagner, S., Jaffé, R., Cawley, K., Dittmar, T. & Stubbins, A. Associations Between the Molecular and Optical Properties of Dissolved Organic Matter in the Florida Everglades, a Model Coastal Wetland System. *Front Chem* **3**, 66 (2015).
331. Lavonen, E. E. *et al.* Tracking changes in the optical properties and molecular composition of dissolved organic matter during drinking water production. *Water Res.* **85**, 286–294 (2015).
332. Timko, S. *et al.* Depth-dependent photodegradation of marine dissolved organic matter. *Frontiers in Marine Science* **2**, (2015).
333. Martínez-Pérez, A. M. *et al.* Linking optical and molecular signatures of dissolved organic matter in the Mediterranean Sea. *Sci. Rep.* **7**, 3436 (2017).
334. Ruckstuhl, K. E., Johnson, E. A. & Miyanishi, K. Introduction. The boreal forest and global change. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 2245–2249 (2008).

335. McKenney, D. W. *et al.* Change and Evolution in the Plant Hardiness Zones of Canada. *Bioscience* **64**, 341–350 (2014).
336. Jaffé, R. *et al.* Dissolved Organic Matter in Headwater Streams: Compositional Variability across Climatic Regions of North America. *Geochim. Cosmochim. Acta* **94**, 95–108 (2012).
337. Goldblum, D. & Rigg, L. S. The Deciduous Forest – Boreal Forest Ecotone. *Geography Compass* **4**, 701–717 (2010).
338. Liu, F. & Wang, D. Dissolved organic carbon concentration and biodegradability across the global rivers: A meta-analysis. *Sci. Total Environ.* **818**, 151828 (2022).
339. Mulholland, P. J. 6 - Large-Scale Patterns in Dissolved Organic Carbon Concentration, Flux, and Sources. in *Aquatic Ecosystems* (eds. Findlay, S. E. G. & Sinsabaugh, R. L.) 139–159 (Academic Press, 2003).
340. Murphy, K. R., Stedmon, C. A., Wenig, P. & Bro, R. OpenFluor– an online spectral library of auto-fluorescence by organic compounds in the environment. *Anal. Methods* **6**, 658–661 (2014).
341. Murphy, K. R., Ruiz, G. M., Dunsmuir, W. T. M. & Waite, T. D. Optimized parameters for fluorescence-based verification of ballast water exchange by ships. *Environ. Sci. Technol.* **40**, 2357–2362 (2006).
342. Fellman, J. B., Hood, E. & Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnol. Oceanogr.* **55**, 2452–2462 (2010).
343. Coble, P. G., Del Castillo, C. E. & Avril, B. Distribution and optical properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **45**, 2195–2223 (1998).
344. Coble, P. G., Green, S. A., Blough, N. V. & Gagosian, R. B. Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy. *Nature* **348**, 432–435 (1990).
345. Parlanti, E., Wörz, K., Geoffroy, L. & Lamotte, M. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. *Org. Geochem.* **31**, 1765–1781 (2000).
346. Ertel, J. R., Hedges, J. I. & Perdue, E. M. Lignin signature of aquatic humic substances. *Science* **223**, 485–487 (1984).

347. Liu, Z., Sleighter, R. L., Zhong, J. & Hatcher, P. G. The chemical changes of DOM from black waters to coastal marine waters by HPLC combined with ultrahigh resolution mass spectrometry. *Estuar. Coast. Shelf Sci.* **92**, 205–216 (2011).
348. D'Andrilli, J., Silverman, V., Buckley, S. & Rosario-Ortiz, F. L. Inferring Ecosystem Function from Dissolved Organic Matter Optical Properties: A Critical Review. *Environ. Sci. Technol.* **56**, 11146–11161 (2022).
349. Cawley, K. M., Campbell, J., Zwilling, M. & Jaffé, R. Evaluation of forest disturbance legacy effects on dissolved organic matter characteristics in streams at the Hubbard Brook Experimental Forest, New Hampshire. *Aquat. Sci.* **76**, 611–622 (2014).
350. Lenoir, J. & Svenning, J.-C. Climate-related range shifts - a global multidimensional synthesis and new research directions. *Ecography* **38**, 15–28 (2015).
351. Yalcin, S. & Leroux, S. J. Diversity and suitability of existing methods and metrics for quantifying species range shifts. *Glob. Ecol. Biogeogr.* **26**, 609–624 (2017).
352. Leithead, M. D., Anand, M. & Silva, L. C. R. Northward migrating trees establish in treefall gaps at the northern limit of the temperate–boreal ecotone, Ontario, Canada. *Oecologia* **164**, 1095–1106 (2010).
353. Lembrechts, J. J. *et al.* Disturbance is the key to plant invasions in cold environments. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14061–14066 (2016).
354. Pamerleau-Couture, É., Rossi, S., Pothier, D. & Krause, C. Wood properties of black spruce (Picea mariana (Mill.) BSP) in relation to ring width and tree height in even- and uneven-aged boreal stands. *Ann. For. Sci.* **76**, 1–13 (2019).
355. Baltzer, J. L. *et al.* Increasing fire and the decline of fire adapted black spruce in the boreal forest. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
356. Česonienė, L. *et al.* Effects of Clear-Cuts in Scots Pine-Dominated Forests on Vaccinium myrtillus and Vaccinium vitis-idaea Vegetative Characteristics, and Accu-mulation of Phenolic Compounds. *Baltic For.* **24**, 278–286 (2018).
357. Yu, H. *et al.* Impact of dataset diversity on accuracy and sensitivity of parallel factor analysis model of dissolved organic matter fluorescence excitation-emission matrix. *Sci. Rep.* **5**, 10207 (2015).

358. Kasischke, E. S. & Stocks, B. J. *Fire, Climate Change, and Carbon Cycling in the Boreal Forest*. (Springer Science & Business Media, 2012).
359. Andrews, C., Foster, J. R., Weiskittel, A., D'Amato, A. W. & Simons-Legaard, E. Integrating historical observations alters projections of eastern North American spruce–fir habitat under climate change. *Ecosphere* **13**, (2022).
360. Zhu, H. *et al.* The Ontario Climate Data Portal, a user-friendly portal of Ontario-specific climate projections. *Sci Data* **7**, 147 (2020).
361. Han, H. *et al.* Acidification impacts on the molecular composition of dissolved organic matter revealed by FT-ICR MS. *Sci. Total Environ.* **805**, 150284 (2022).
362. Pace, M. L. *et al.* pH change induces shifts in the size and light absorption of dissolved organic matter. *Biogeochemistry* **108**, 109–118 (2012).
363. Simpson, A. J. *et al.* Molecular structures and associations of humic substances in the terrestrial environment. *Naturwissenschaften* **89**, 84–88 (2002).
364. Derrien, M., Lee, Y. K., Shin, K.-H. & Hur, J. Comparing discrimination capabilities of fluorescence spectroscopy versus FT-ICR-MS for sources and hydrophobicity of sediment organic matter. *Environ. Sci. Pollut. Res. Int.* **25**, 1892–1902 (2018).
365. Hudson, D. T., Leach, J. A., Webster, K. & Houle, D. Streamflow regime of a lake-stream system based on long-term data from a high-density hydrometric network. *Hydrol. Process.* **35**, (2021).
366. Charbonneau, K. L. *et al.* Are There Longitudinal Effects of Forest Harvesting on Carbon Quality and Flow and Methylmercury Bioaccumulation in Primary Consumers of Temperate Stream Networks? *Environ. Toxicol. Chem.* **41**, 1490–1507 (2022).
367. Sidhu. This paper is in preparation.
368. Erdozain, M. *et al.* Forest management impacts on stream integrity at varying intensities and spatial scales: Do abiotic effects accumulate spatially? *Sci. Total Environ.* **753**, 141968 (2021).
369. Roux, M. Basic Procedures in Hierarchical Cluster Analysis. in *Applied Multivariate Analysis in SAR and Environmental Studies* (eds. Devillers, J. & Karcher, W.) 115–135 (Springer Netherlands, 1991).

370. Murtagh, F. & Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J. Classification* **31**, 274–295 (2014).
371. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *arXiv [cs.LG]* 2825–2830 (2012).
372. Lindsay, J. B. WhiteboxTools user manual. *Geomorphometry and Hydrogeomatics Research Group, University of Guelph, Guelph, Canada* **20**, (2018).
373. Ågren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J. & Arp, P. A. Evaluating digital terrain indices for soil wetness mapping – a Swedish case study. *Hydrol. Earth Syst. Sci.* **18**, 3623–3634 (2014).
374. Pike, R. J., Evans, I. S. & Hengl, T. Chapter 1 Geomorphometry: A Brief Guide. in *Developments in Soil Science* (eds. Hengl, T. & Reuter, H. I.) vol. 33 3–30 (Elsevier, 2009).
375. Wilson, J. P., Gallant, J. C. & Others. Digital terrain analysis. *Terrain analysis: Principles and applications* **6**, 1–27 (2000).
376. Beven, K. J. & Kirkby, M. J. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol. Sci. Bull.* **24**, 43–69 (1979).
377. Horn, B. K. P. Hill shading and the reflectance map. *Proc. IEEE* **69**, 14–47 (1981).
378. Murphy, P. N. C. *et al.* Modelling and mapping topographic variations in forest soils at high resolution: A case study. *Ecol. Modell.* **222**, 2314–2332 (2011).
379. Buttle, J. M. The Effects of Forest Harvesting on Forest Hydrology and Biogeochemistry. in *Forest Hydrology and Biogeochemistry: Synthesis of Past Research and Future Directions* (eds. Levia, D. F., Carlyle-Moses, D. & Tanaka, T.) 659–677 (Springer Netherlands, 2011).
380. Murphy, P. N. C., Ogilvie, J. & Arp, P. Topographic modelling of soil moisture conditions: a comparison and verification of two models. *Eur. J. Soil Sci.* **60**, 94–109 (2009).
381. Moore, I. D. *Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications*. (Water Research Foundation of Australia, 1991).

382. Ågren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J. & Arp, P. A. Evaluating digital terrain indices for soil wetness mapping--a Swedish case study. *Hydrol. Earth Syst. Sci.* **18**, 3623–3634 (2014).
383. Bilyk, A., Pulkki, R., Shahi, C. & Larocque, G. R. Development of the Ontario Forest Resources Inventory: a historical review. *Can. J. For. Res.* **51**, 198–209 (2021).
384. Yang, Q., Meng, F.-R., Bourque, C. P.-A. & Zhao, Z. Production of high-resolution forest-ecosite maps based on model predictions of soil moisture and nutrient regimes over a large forested area. *Sci. Rep.* **7**, 10998 (2017).
385. Bose, A. K., Harvey, B. D., Brais, S., Beaudet, M. & Leduc, A. Constraints to partial cutting in the boreal forest of Canada in the context of natural disturbance-based management: a review. *Forestry* **87**, 11–28 (2013).
386. Heeney, C. J. & Ontario. Forest Management Branch. *A Silvicultural Guide to the Aspen Working Group in Ontario*. (Forest Management Branch, Forest Research Branch, Division of Forests, Ontario Ministry of Natural Resources, 1975).
387. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 307 (2008).
388. Genuer, R., Poggi, J.-M. & Tuleau, C. Random Forests: some methodological insights. *arXiv [stat.ML]* (2008).
389. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
390. Lechtenfeld, O. J., Hertkorn, N., Shen, Y., Witt, M. & Benner, R. Marine sequestration of carbon in bacterial metabolites. *Nature Communications* vol. 6 Preprint at https://doi.org/10.1038/ncomms7711 (2015).
391. Hach, P. F. *et al.* Rapid microbial diversification of dissolved organic matter in oceanic surface waters leads to carbon sequestration. *Sci. Rep.* **10**, 13025 (2020).
392. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences* **103**, 12115–12120 (2006).
393. Mentges, A. *et al.* Microbial Physiology Governs the Oceanic Distribution of Dissolved Organic Carbon in a Scenario of Equal Degradability. *Frontiers in Marine Science* **7**, (2020).

394. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016).
395. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
396. Ferenci, T. Trade-off Mechanisms Shaping the Diversity of Bacteria. *Trends Microbiol.* **24**, 209–223 (2016).
397. Banerjee, S., Schlaeppi, K. & van der Heijden, M. G. A. Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microbiol.* **16**, 567–576 (2018).
398. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).
399. Proulx, S. R., Promislow, D. E. L. & Phillips, P. C. Network thinking in ecology and evolution. *Trends Ecol. Evol.* **20**, 345–353 (2005).
400. Mills, M. B. *et al.* Tropical forests post-logging are a persistent net carbon source to the atmosphere. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2214462120 (2023).
401. Mostovaya, A., Hawkes, J. A., Koehler, B., Dittmar, T. & Tranvik, L. J. Emergence of the Reactivity Continuum of Organic Matter from Kinetics of a Multitude of Individual Molecular Constituents. *Environ. Sci. Technol.* **51**, 11571–11579 (2017).
402. Arnosti, C., Steen, A. D., Ziervogel, K., Ghobrial, S. & Jeffrey, W. H. Latitudinal gradients in degradation of marine dissolved organic carbon. *PLoS One* **6**, e28900 (2011).
403. Carlson, C. A. *et al.* Interactions among dissolved organic carbon, microbial processes, and community structure in the mesopelagic zone of the northwestern Sargasso Sea. *Limnol. Oceanogr.* **49**, 1073–1083 (2004).
404. Shen, Y. & Benner, R. Mixing it up in the ocean carbon cycle and the removal of refractory dissolved organic carbon. *Sci. Rep.* **8**, 2542 (2018).
405. Hertkorn, N., Harir, M., Koch, B. P., Michalke, B. & Schmitt-Kopplin, P. High-field NMR spectroscopy and FTICR mass spectrometry: powerful discovery tools for the molecular level characterization of marine dissolved organic matter. *Biogeosciences* **10**, 1583–1624 (2013).

406. Ball, G. I. & Aluwihare, L. I. CuO-oxidized dissolved organic matter (DOM) investigated with comprehensive two dimensional gas chromatography-time of flight-mass spectrometry (GC×GC-TOF-MS). *Org. Geochem.* **75**, 87–98 (2014).
407. Becker, S. *et al.* Laminarin is a major molecule in the marine carbon cycle. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 6599–6607 (2020).
408. Mentges, A., Feenders, C., Deutsch, C., Blasius, B. & Dittmar, T. Long-term stability of marine dissolved organic carbon emerges from a neutral network of compounds and microbes. *Sci. Rep.* **9**, 17780 (2019).
409. Behnke, M. I. *et al.* Pan-arctic riverine dissolved organic matter: Synchronous molecular stability, shifting sources and subsidies. *Global Biogeochem. Cycles* **35**, (2021).
410. Barton. MuMIn: multi-model inference. R package version 1. 0. 0. *http://r-forge. r-project. org/projects/mumin/*.
411. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
412. Hutchins, R. H. S. *et al.* The optical, chemical, and molecular dissolved organic matter succession along a boreal soil-stream-river continuum. *J. Geophys. Res. Biogeosci.* **122**, 2892–2908 (2017).
413. Menkis, A. *et al.* Occurrence and impact of the root-rot biocontrol agent Phlebiopsis gigantea on soil fungal communities in Picea abies forests of northern Europe. *FEMS Microbiol. Ecol.* **81**, 438–445 (2012).
414. White, T. J., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. in *PCR Protocols* 315–322 (Elsevier, 1990).
415. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
416. Nakagawa, S., Johnson, P. C. D. & Schielzeth, H. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* **14**, (2017).
417. McKnight, D. M. *et al.* Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnol. Oceanogr.* **46**, 38–48 (2001).

418. Huguet, A. *et al.* Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Org. Geochem.* **40**, 706–719 (2009).
419. Ågren, A. *et al.* Dissolved organic carbon characteristics in boreal streams in a forest-wetland gradient during the transition between winter and summer. *Journal of Geophysical Research* vol. 113 Preprint at https://doi.org/10.1029/2007jg000674 (2008).
420. Hoekstra, J. M. *et al. The Atlas of Global Conservation: Changes, Challenges and Opportunities to Make a Difference*. (University of California Press, 2010).
421. Gullian-Klanian, M., Gold-Bouchot, G., Delgadillo-Díaz, M., Aranda, J. & Sánchez-Solís, M. J. Effect of the use of Bacillus spp. on the characteristics of dissolved fluorescent organic matter and the phytochemical quality of Stevia rebaudiana grown in a recirculating aquaponic system. *Environ. Sci. Pollut. Res. Int.* **28**, 36326–36343 (2021).
422. Chai, L. *et al.* Urbanization altered regional soil organic matter quantity and quality: Insight from excitation emission matrix (EEM) and parallel factor analysis (PARAFAC). *Chemosphere* **220**, 249–258 (2019).
423. Zhou, Y., Martin, P. & Müller, M. Composition and cycling of dissolved organic matter from tropical peatlands of coastal Sarawak, Borneo, revealed by fluorescence spectroscopy and parallel factor analysis. *Biogeosciences* **16**, 2733–2749 (2019).
424. Lambert, T. *et al.* Effects of human land use on the terrestrial and aquatic sources of fluvial organic matter in a temperate river basin (The Meuse River, Belgium). *Biogeochemistry* **136**, 191–211 (2017).
425. Søndergaard, M., Stedmon, C. A. & Borch, N. H. Fate of terrigenous dissolved organic matter (DOM) in estuaries: Aggregation and bioavailability. *Ophelia* **57**, 161–176 (2003).
426. Osburn, C. L., Wigdahl, C. R., Fritz, S. C. & Saros, J. E. Dissolved organic matter composition and photoreactivity in prairie lakes of the U.S. Great Plains. *Limnol. Oceanogr.* **56**, 2371–2390 (2011).
427. Kothawala, D. N., von Wachenfeldt, E., Koehler, B. & Tranvik, L. J. Selective loss and preservation of lake water dissolved organic matter fluorescence during long-term dark incubations. *Sci. Total Environ.* **433**, 238–246 (2012).

428. Andersson, M. G. I., Catalán, N., Rahman, Z., Tranvik, L. J. & Lindström, E. S. Effects of sterilization on dissolved organic carbon (DOC) composition and bacterial utilization of DOC from lakes. *Aquat. Microb. Ecol.* **82**, 199–208 (2018).
429. Imbeau, E., Vincent, W. F., Wauthy, M., Cusson, M. & Rautio, M. Hidden stores of organic matter in northern lake ice: Selective retention of terrestrial particles, phytoplankton and labile carbon. *J. Geophys. Res. Biogeosci.* **126**, (2021).
430. Jutaporn, P., Muenphukhiaw, N., Phungsai, P., Leungprasert, S. & Musikavong, C. Characterization of DBP precursor removal by magnetic ion exchange resin using spectroscopy and high-resolution mass spectrometry. *Water Res.* **217**, 118435 (2022).
431. Yamashita, Y., Scinto, L. J., Maie, N. & Jaffé, R. Dissolved Organic Matter Characteristics Across a Subtropical Wetland's Landscape: Application of Optical Properties in the Assessment of Environmental Dynamics. *Ecosystems* **13**, 1006–1019 (2010).
432. Dainard, P. G., Guéguen, C., McDonald, N. & Williams, W. J. Photobleaching of fluorescent dissolved organic matter in Beaufort Sea and North Atlantic Subtropical Gyre. *Mar. Chem.* **177**, 630–637 (2015).
433. Cawley, K. M. *et al.* Identifying fluorescent pulp mill effluent in the Gulf of Maine and its watershed. *Mar. Pollut. Bull.* **64**, 1678–1687 (2012).
434. Yang, L., Chen, Y., Lei, J. & Zhu, Z. Effects of coastal aquaculture on sediment organic matter: Assessed with multiple spectral and isotopic indices. *Water Res.* **223**, 118951 (2022).
435. Wasswa, J., Driscoll, C. T. & Zeng, T. Contrasting Impacts of Photochemical and Microbial Processing on the Photoreactivity of Dissolved Organic Matter in an Adirondack Lake Watershed. *Environ. Sci. Technol.* **56**, 1688–1701 (2022).
436. Wang, S., Perkins, M., Matthews, D. A. & Zeng, T. Coupling Suspect and Nontarget Screening with Mass Balance Modeling to Characterize Organic Micropollutants in the Onondaga Lake-Three Rivers System. *Environ. Sci. Technol.* **55**, 15215–15226 (2021).
437. Coulson, L. E. *et al.* Small rain events during drought alter sediment dissolved organic carbon leaching and respiration in intermittent stream sediments. *Biogeochemistry* **159**, 159–178 (2022).

438. Wauthy, M. *et al.* Increasing dominance of terrigenous organic matter in circumpolar freshwaters due to permafrost thaw. *Limnol. Oceanogr. Lett.* **3**, 186–198 (2018).
439. Yamashita, Y., Kojima, D., Yoshida, N. & Shibata, H. Relationships between dissolved black carbon and dissolved organic matter in streams. *Chemosphere* **271**, 129824 (2021).
440. Wang, S. *et al.* Organic Micropollutants in New York Lakes: A Statewide Citizen Science Occurrence Study. *Environ. Sci. Technol.* **54**, 13759–13770 (2020).
441. Zhuang, W.-E., Chen, W., Cheng, Q. & Yang, L. Assessing the priming effect of dissolved organic matter from typical sources using fluorescence EEMs-PARAFAC. *Chemosphere* **264**, 128600 (2021).
442. Groeneveld, M. *et al.* Selective adsorption of terrestrial dissolved organic matter to inorganic surfaces along a boreal inland water continuum. *J. Geophys. Res. Biogeosci.* **125**, (2020).
443. Eder, A. *et al.* Pathways and composition of dissolved organic carbon in a small agricultural catchment during base flow conditions. *Ecohydrology & Hydrobiology* **22**, 96–112 (2022).
444. Wang, H. *et al.* Spectral and isotopic characteristics of particulate organic matter in a subtropical estuary under the influences of human disturbance. *J. Mar. Syst.* **203**, 103264 (2020).
445. Lin, H. & Guo, L. Variations in Colloidal DOM Composition with Molecular Weight within Individual Water Samples as Characterized by Flow Field-Flow Fractionation and EEM-PARAFAC Analysis. *Environ. Sci. Technol.* **54**, 1657–1667 (2020).
446. Brym, A. *et al.* Optical and chemical characterization of base-extracted particulate organic matter in coastal marine environments. *Mar. Chem.* **162**, 96–113 (2014).
447. Heibati, M. *et al.* Assessment of drinking water quality at the tap using fluorescence spectroscopy. *Water Res.* **125**, 1–10 (2017).
448. Kida, M. *et al.* Origin, distributions, and environmental significance of ubiquitous humic-like fluorophores in Antarctic lakes and streams. *Water Res.* **163**, 114901 (2019).

449. Yang, L., Cheng, Q., Zhuang, W.-E., Wang, H. & Chen, W. Seasonal changes in the chemical composition and reactivity of dissolved organic matter at the land-ocean interface of a subtropical river. *Environ. Sci. Pollut. Res. Int.* **26**, 24595–24608 (2019).
450. Yamashita, Y., Maie, N., Briceño, H. & Jaffé, R. Optical characterization of dissolved organic matter in tropical rivers of the Guayana Shield, Venezuela. *Journal of Geophysical Research* vol. 115 Preprint at https://doi.org/10.1029/2009jg000987 (2010).
451. Walker, S. A., Amon, R. M. W. & Stedmon, C. The use of PARAFAC modeling to trace terrestrial dissolved organic matter and fingerprint water masses in coastal Canadian Arctic surface waters. *Journal of* doi:10.1029/2009JG000990.
452. Stedmon, C. A. *et al.* Characteristics of dissolved organic matter in Baltic coastal sea ice: allochthonous or autochthonous origins? *Environ. Sci. Technol.* **41**, 7273–7279 (2007).
453. Jaffé, R., Cawley, K. M. & Yamashita, Y. Applications of Excitation Emission Matrix Fluorescence with Parallel Factor Analysis (EEM-PARAFAC) in Assessing Environmental Dynamics of Natural Dissolved Organic Matter (DOM) in Aquatic Environments: A Review. *ACS Symposium Series* 27–73 Preprint at https://doi.org/10.1021/bk-2014-1160.ch003 (2014).
454. Lambert, T. *et al.* Along-stream transport and transformation of dissolved organic matter in a large tropical river. *Biogeosciences* **13**, 2727–2741 (2016).
455. Graeber, D., Gelbrecht, J., Pusch, M. T., Anlanger, C. & von Schiller, D. Agriculture has changed the amount and composition of dissolved organic matter in Central European headwater streams. *Sci. Total Environ.* **438**, 435–446 (2012).
456. Marcé, R., Verdura, L. & Leung, N. Dissolved organic matter spectroscopy reveals a hot spot of organic matter changes at the river–reservoir boundary. *Aquat. Sci.* **83**, 67 (2021).
457. Amaral, V., Romera-Castillo, C. & Forja, J. Submarine mud volcanoes as a source of chromophoric dissolved organic matter to the deep waters of the Gulf of Cádiz. *Sci. Rep.* **11**, 3200 (2021).

458. Schittich, A.-R. *et al.* Investigating Fluorescent Organic-Matter Composition as a Key Predictor for Arsenic Mobility in Groundwater Aquifers. *Environ. Sci. Technol.* **52**, 13027–13036 (2018).
459. Kulkarni, H. V., Mladenov, N., Johannesson, K. H. & Datta, S. Contrasting dissolved organic matter quality in groundwater in Holocene and Pleistocene aquifers and implications for influencing arsenic mobility. *Appl. Geochem.* **77**, 194–205 (2017).
460. Asmala, E. *et al.* Eutrophication leads to accumulation of recalcitrant autochthonous organic matter in coastal environment. *Global Biogeochem. Cycles* **32**, 1673–1687 (2018).
461. Kowalczuk, P. *et al.* Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: Interannual variability. *Mar. Chem.* **113**, 182–196 (2009).
462. Hambly, A. C. *et al.* Characterising organic matter in recirculating aquaculture systems with fluorescence EEM spectroscopy. *Water Res.* **83**, 112–120 (2015).
463. DeFrancesco, C. & Guéguen, C. Long-term trends in dissolved organic matter composition and its relation to sea ice in the Canada basin, arctic ocean (2007–2017). *J. Geophys. Res. C: Oceans* **126**, (2021).
464. Queimaliños, C. *et al.* Linking landscape heterogeneity with lake dissolved organic matter properties assessed through absorbance and fluorescence spectroscopy: Spatial and seasonal patterns in temperate lakes of Southern Andes (Patagonia, Argentina). *Sci. Total Environ.* **686**, 223–235 (2019).
465. Wünsch, U. J. *et al.* Quantifying the impact of solid-phase extraction on chromophoric dissolved organic matter composition. *Mar. Chem.* **207**, 33–41 (2018).
466. Lambert, T., Bouillon, S., Darchambeau, F., Massicotte, P. & Borges, A. V. Shift in the chemical composition of dissolved organic matter in the Congo River network. *Biogeosciences* **13**, 5405–5420 (2016).
467. Yamashita, Y., Kloeppel, B. D., Knoepp, J., Zausen, G. L. & Jaffé, R. Effects of Watershed History on Dissolved Organic Matter Characteristics in Headwater Streams. *Ecosystems* **14**, 1110–1122 (2011).

468. Osburn, C. L. & Stedmon, C. A. Linking the chemical and optical properties of dissolved organic matter in the Baltic–North Sea transition zone to differentiate three allochthonous inputs. *Mar. Chem.* **126**, 281–294 (2011).
469. Stedmon, C. A. & Markager, S. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnol. Oceanogr.* **50**, 686–697 (2005).
470. Shutova, Y., Baker, A., Bridgeman, J. & Henderson, R. K. Spectroscopic characterisation of dissolved organic matter changes in drinking water treatment: From PARAFAC analysis to online monitoring wavelengths. *Water Res.* **54**, 159–169 (2014).
471. Cawley, K. M., Ding, Y., Fourqurean, J. & Jaffé, R. Characterising the sources and fate of dissolved organic matter in Shark Bay, Australia: a preliminary study using optical properties and stable carbon isotopes. *Mar. Freshwater Res.* **63**, 1098–1107 (2012).
472. Smith, M. A., Kominoski, J. S., Gaiser, E. E., Price, R. M. & Troxler, T. G. Stormwater runoff and tidal flooding transform dissolved organic matter composition and increase bioavailability in urban coastal ecosystems. *J. Geophys. Res. Biogeosci.* **126**, (2021).
473. Chen, M., Jung, J., Lee, Y. K. & Hur, J. Surface accumulation of low molecular weight dissolved organic matter in surface waters and horizontal off-shelf spreading of nutrients and humic-like fluorescence in the Chukchi Sea of the Arctic Ocean. *Sci. Total Environ.* **639**, 624–632 (2018).
474. Retelletti Brogi, S., Jung, J. Y., Ha, S.-Y. & Hur, J. Seasonal differences in dissolved organic matter properties and sources in an Arctic fjord: Implications for future conditions. *Sci. Total Environ.* **694**, 133740 (2019).
475. D'Andrilli, J., Junker, J. R., Smith, H. J., Scholl, E. A. & Foreman, C. M. DOM composition alters ecosystem function during microbial processing of isolated sources. *Biogeochemistry* **142**, 281–298 (2019).
476. Sun, Y. *et al.* Hillslopes in headwaters of Qinghai-Tibetan plateau as hotspots for subsurface dissolved organic carbon processing during permafrost thaw. *J. Geophys. Res. Biogeosci.* **126**, (2021).