Terrestrial and coloured dissolved organic matter in Arctic waters: Towards in-situ sensor based monitoring of Arctic-Atlantic organic carbon exchange at major Arctic gateways

Jensen, Anders Dalhoff Bruhn

Publication date:
2023

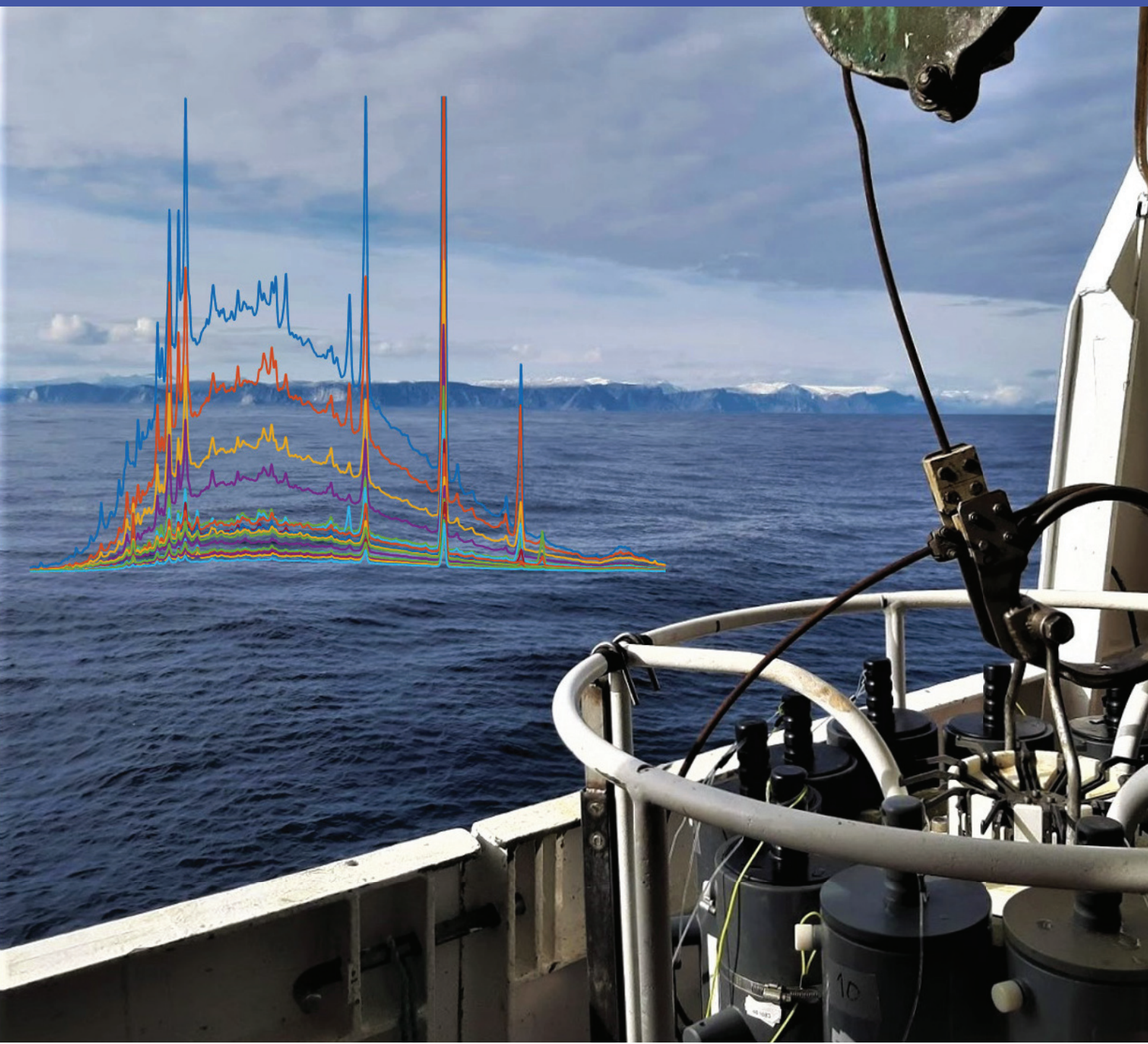
Document Version
Publisher's PDF, also known as Version of record

DTU Aqua
National Institute of Aquatic Resources

Terrestrial and coloured dissolved organic matter in Arctic waters: Towards in-situ sensor based monitoring of Arctic-Atlantic organic carbon exchange at major Arctic gateways

Anders Dalhoff Bruhn

PhD Thesis



Ph.D. Thesis

Doctor in Philosophy

Terrestrial and coloured dissolved organic matter in Arctic waters: Towards in-situ sensor based monitoring of Arctic-Atlantic organic carbon exchange at major Arctic gateways

Anders Dalhoff Bruhn

Supervisor:

Colin Andrew Stedmon

Co-supervisor:

Christopher Lee Osburn

DTU Aqua 2023

PREFACE

The research presented in the Ph.D. thesis has been conducted to fulfil the requirements of acquiring a Doctor degree in Philosophy. The majority of the research was conducted at the National Institute of Aquatic Resources at the Technical University of Denmark (DTU Aqua) between September 2019 until September 2022 at the Section for Oceans and Arctic under the supervision of main supervisor Professor Colin A. Stedmon. In February and March 2022 a part of the research was conducted at North Carolina State University (NCSU) in the United States under the supervision of co-supervisor Professor Christopher L. Osburn. The research was funded by Independent Research Fund Denmark (No. 9040-938 00266B). The external stay to NSCU was supported by the Kaj og Hermilla Ostenfeldt Fund.

The Ph.D. thesis is composed of a synopsis followed by three chapters in the form of independent scientific papers (Paper A, B and C). The synopsis first provides an introduction, which serves the reader with an overview of the scientific landscape including fundamental concepts, results from other studies, methodologies and current research gaps. The synopsis then follows up with of the most important findings from the three chapters together with a discussion of the work and a perspective on how these findings contribute to the state-of-the-art. Each of the scientific papers are structured according to common journal styles, starting with an individual introduction, then description of methodologies and laboratory work, followed by results including figures and tables, and finally broad discussions leading to conclusion on the findings.

"Even small changes in the DOC pool in the ocean can lead to disturbance of the global carbon cycle on 1000 to 10.000 years scale" – John Hedges, 2002

DTU Aqua, Kongens Lyngby, 5th March 2023

Anders Dalhoff Bruhn

ACKNOWLEDGEMENT

Working with the ocean has always been a dream of mine. Combining it with research was something I rarely got introduced to during my study time. Becoming a chemical oceanographer was something I never thought of. However, starting at DTU Aqua five years ago, as an assistant, made it possible for me to meet a group of inspiring people, who I owe a lot of the reason for why I am here today.

The person most responsible for opening my eyes to the world of chemical oceanography is my supervisor Colin Stedmon. I clearly remember when I was put in front of a fluorometer and starred at measurements of dissolved organic matter in the ocean for the first time. I was surprised about how much knowledge you had and the amount of information you could extract from such a “simple” measurement. However, at the same time I was also baffled about how much still needed to be uncovered. That was the mystery which really ignited my interest for diving into this field and try to uncover the unknown. However, without your trust in me to do the task, I am not sure I would have started this journey at all. Whenever I have had a question during my Ph.D. study you have always had your door open and always found time to discuss. I admire you for always listening carefully when I have presented new thoughts and ideas and tried to test the field. The trust from you to carry out my ideas is what have made many of the results possible. It is probably also the main reason why I today sit here with a Ph.D. thesis in my hands. It is a trust I rarely find from anyone. I want to thank you immensely for that.

The next person in line is Urban Wünsch, currently post doc at DTU Aqua. You taught me a good lesson one time and it will forever stick with me: “Work smart. Not hard.” When I started my Ph.D. study I knew very little about coding and various type of chemometrics. Without your guidance through some of the struggles with the simplest and most complex matters, I would probably still be writing my first line of code in this moment. Besides from being my guide, you have also been the one I could talk to about anything, always finding your time for inspiring discussions – often one single question became an hour of turning the entire science field up and down. At time most needed, you have always came to my aid. Even though you have not been my official co-supervisor, you definitely earned that title.

I want to thank my co-supervisor Christopher Osburn, who I visited for my external stay at North Carolina State University. You made me feel most welcome and at times almost at home under your wings. It was a great experience to visit your lab group and from our conversations, I brought home a lot of new knowledge. Your excitement in my project and the discoveries I made underway, even the smallest ones, kept me motivated and made me feel I was on the right track. The time in North Carolina still stands as one of my best memories from the Ph.D. journey.

A special thanks must be given to Rafael Gonçalves-Araujo, currently post doc at DTU Aqua, but now also a dear friend of mine. Thanks for being the person to always cheer me up and keep sure I was staying in touch with myself during this whole journey. You were in charge of coordinating the cruise to the Fram Strait (ARICE-NoTAC project) and was the one who made sure I got the samples needed to create the results I have today. We have sailed the Arctic

Ocean together and I would not have wished to do it with anyone else – and I will do it all over again if I ever get the chance. Keep rocking the world!

I also want to thank Johanna Sjöstedt, researcher at Lund University, who I collaborated with on the first paper. With you, I have shared many joyful moments inside and outside the laboratory, always getting the best out of any kind of situation. It was fantastic to be a part of your team at Lund University and I always felt appreciated with you as a host. Definitely a time I will never forget. I sure miss our times working together and hopefully there will be another chance in the future.

I would like to thank all the people at DTU Aqua who I have crossed paths with, Ph.D. students, professors, technicians, secretaries and master students, who have all helped me and inspired me in some way along this journey.

Then I would like to thank my family for their patience and understanding whenever the Ph.D. study needed to be prioritized. They have always reminded me where I came from and to stay true to myself. At the same time they have taught me to never give up and follow my dreams. My mother has never doubted me a single time – and I love you for that.

Lastly, but most importantly, I would like to thank my fantastic life partner Sandra, who has always been my solid rock. You have always been there to support and cheer me up, even when I have been in the darkest corners of the Ph.D. study, and somehow you have always managed to drag me back to see the light. You have reminded me to always be proud of myself and the things I achieved along the way, and we have celebrated moments I could not share with anyone else. But to be honest, I am proud of you. You have giving me the time to devote myself completely to this Ph.D. study, with the fullest of my heart, often on behalf of my time with you. For that I will always love you.

SUMMARY

With accelerating global warming, air temperatures around the globe are increasing and the highest increase has been observed over the past decades, particularly in the Arctic region. An increase in air temperatures for the Arctic region will not only affect the rate of sea ice melting, but will also increase river discharge and erosion of riverbanks and coastlines. The latter processes will lead to more terrestrial dissolved organic matter (tDOM) being transported from land to the Arctic Ocean in the future. The fate of dissolved organic carbon (DOC) supplied as tDOM is of great concern, since it potentially can be transformed to carbon dioxide (CO_2) and be exchanged with the atmosphere.

The aim of this Ph.D. thesis was to study the fate of tDOM in the Arctic Ocean and to develop an analytical method to quantify and trace its distribution from different sources. To understand the continuum of tDOM, the fate of the dissolved material was first investigated in the Arctic coastal zone (ACZ), where it is released, and later in Arctic major gateway, the Fram Strait, where it is exported to the Atlantic Ocean.

My studies found that the fate of tDOM from coastal erosion depends on the permafrost soil type being dissolved into the coastal waters. The derived tDOM from three different permafrost types led to the development of three distinct marine bacterial communities, which also led to three different bacterial growth efficiencies. The difference in bacterial growth efficiency ultimately means that the carbon processing of the derived DOC, and the remineralization to CO_2, from coastal erosion is greatly dependent on which permafrost soil type will erode into coastal waters. However, it was found that most of the DOC was likely refractory to rapid mineralization and may survive passage through the coastal zone.

The refractory part of tDOM that survives the coastal zone will be exported to the open ocean, where it will follow ocean currents across the Arctic Ocean and under the sea ice. The water masses circulating in the Arctic Ocean will eventually get exported through one of two major Arctic gateways, either side of Greenland, to the Atlantic Ocean. Once in the Atlantic Ocean, the exported tDOM and associated DOC may be sequestered into the deep ocean and despite eventual mineralization with time, be kept away from interaction with the atmosphere. To follow how the carbon cycle is responding to climate change it is therefore important to trace this pathway. One approach to obtain a quantitative measure of the distribution and fate of tDOM, is to use the biopolymer lignin as a biomarker. Lignin only exists in terrestrial plants and when found in the ocean reflects tDOM distribution. Besides that, lignin character has also been shown to be able to provide information on source and diagenesis of the lignin material.

During this project, I developed a machine-learning assisted method to quantify lignin phenols in seawater with the aim of improving the sensitivity and specificity of high-pressure liquid chromatography (HPLC) coupled with absorbance detection. This new method circumnavigates current limitations with this instrumentation and substantially reduces seawater needed for measurement. The method was applied to seawater from the Fram Strait and the total concentration of dissolved lignin was clearly higher for Arctic surface waters being exported, than Atlantic waters being imported, reflecting the tDOM supplied from landmasses around the Arctic Ocean and persisting to reach the Fram Strait and thereby exported to the

North Atlantic. Additionally it was found that multiple sources of terrestrial material could be differentiated in the exported water masses.

Finally, I investigated the relationship between DOM fluorescence and lignin phenol concentrations in seawater using N-way Partial Least Squares (N-PLS) regression. The goal was to predict the essential lignin parameters from spectral fluorescence measurements which are much less time consuming to make and for which there is a time series of data for in the Fram Strait. The N-PLS model derived successfully replicated the measured trends in the water masses sampled and could be used to predict lignin phenol ratios and thereby differentiate between tDOM sources. The developed N-PLS model was further reduced to predict lignin phenols based on only four excitation wavelengths in order to investigate the potential of designing in situ sensors for the purpose. Despite the large reduction in excitations wavelengths, the model still performed well. This indicates that the approach holds promise as a proxy for estimating lignin concentrations, greatly extending potential spatial and temporal coverage, and paving the way for development of sensors which can be used on profiling (automated) systems and help monitor and quantify the effect of climate change on the ocean's carbon budget more closely in the future.

RESÚME

I en verden med accelererende global opvarmning stiger lufttemperaturerne rundt omkring på kloden, og især i den Arktiske region er den højeste stigning blevet observeret. Stigninger i lufttemperaturen i Arktis vil ikke kun påvirke smeltningen af havisens, men vil også føre til en øget mængde af vand i floderne og intensiveret erosion af flodbredder og kystlinjer rundt omkring i det Arktiske Hav. Sidstnævnte konsekvenser vil endvidere føre til at mere terrestrisk opløst organisk stof (tDOM) transporteres fra land ud i det Arktiske Hav i fremtiden. En stor del af tDOM består af kulstof, mere præcist opløst organisk kulstof (DOC). Bekymringen er stor for skæbnen for denne DOC, da den potentielt kan omdannes til CO_2 og derved eksporteres til atmosfæren.

Formålet med Ph.D. afhandlingen var derfor at studere tDOMs skæbne i det Arktiske Hav og udvikle en analytisk metode til at kvantificere og spore dets udbredelse fra forskellige kilder. For at forstå kontinuummet af tDOM i det Arktiske Hav, blev det opløste materiales skæbne først undersøgt i den Arktiske kystzone, hvor det frigives, og senere i ved dybvandsporten, Fram Strædet, hvor det eksporteres til Atlanterhavet.

Ph.D. afhandlingen viser at tDOMs skæbne fra kysterosion afhænger af den type permafrostjord som bliver opløst i kystzonen. Terrestrisk opløst materiale fra tre forskellige permafrosttyper førte til udviklingen af tre forskellige marine bakteriesamfund, da bakterievæksteffektiviteten varierede mellem de forskellige tDOM typer. Dette betyder i sidste ende, at kulstofsbearbejdningen af DOC og mineraliseringen til CO_2 fra kysterosion i høj grad afhænger af hvilken permafrostjordstype, der vil erodere ned i kystzonen. Det viste sig dog, at det meste af DOC faktisk var resistent mod omgående mikrobiologisk nedbrydning og kan derved formentlig overleve kystzonen.

Den resistente del af tDOM, der overlever kystzonen, vil blive eksporteret til det åbne hav, hvor den vil følge havstrømmene rundt i det Arktiske Hav, der løber under havisen. Vandmasserne, der cirkulerer i det Arktiske Hav, vil i sidste ende blive eksporteret videre ud i Atlanterhavet. I Atlanterhavet kan den eksporterede tDOM og DOC blive begravet i de dybere oceaner og derved blive fjernet fra omdannelse til CO_2 i århundreder. Det er derfor afgørende at finde ud af, hvor meget tDOM der vil blive eksporteret. Lignin har vist sig at være en relevant biomarkør til at måle tDOM kvantitativt, da stoffet kun findes i landlevende planter, og derfor afspejler tDOM, hvis det måles i havet. Ud over det har lignin også vist sig at kunne give information om kilden af det terrestriske materiale og nogen af de processer lignin har været udsat for undervejs.

I Ph.D. afhandling bidrager jeg med en machine-learning assisteret metode til at kvantificere lignin phenoler i havvand, hvilket forbedrer følsomheden og specificiteten af højtryksvæskekromatografi (HPLC) kombineret med absorbansdetektion. Den nye metode undgår nuværende begrænsninger for fornævnte instrument og reducerer mængden af havvand der skal indsamles. Den nye metode blev anvendt på havvand fra Fram Strædet, og den samlede koncentration af opløst lignin viste sig at være højere for Arktisk overfladevand, der eksporteres til Atlanterhavet, end Atlanterhavsvand, der importeres til det Arktiske Hav. Samtidig viser det sig at den tDOM der frigives fra landmasser omkring det Arktiske Hav når

hele vejen til Fram Strædet og derfor højst sandssynligt vil blive eksporteret til dybhavet i Atlanterhavet. Derudover blev det også vist at det terrestriske materiale i de eksporterede vandmasser stammer fra forskellige Arktiske floder.

I min Ph.D. afhandling bridrager jerg også med et værktøj til at bestemme koncentrationer af lignin parametre i havvand ud fra fluorescensmålinger baseret på Multilinear Partial Least Square (N-PLS). N-PLS-modellen er den første af sin slags til at bestemme lignin paramtre ud fra spektroskopisk data. N-PLS blev med succes anvendt på fluorescensmålinger på tværs af Fram Strædet og viste sig at afspejle de målte tendenser. Modellen kan bruges til at bestemme lignin phenol sammensætning og derved differentiere mellem kilder af terrestristisk materiale i havet. N-PLS-modellen tilbyder derfor et hurtigt og økonomisk alternativ til HPLC metoden, som samtidig kan give mulighed for en højere prøvegennemstrømning og derved bane vej for større afdækning af områder og over længere tidshorisont i fremtiden. N-PLS modellen blev yderligere reduceret til at bestemme lignin phenoler baseret på fire excitationsbølgelængde. Den reduceret model viste sig at fungere lige så godt som den oprindelige model. Den reducerede N-PLS model kan højst sandsynlig anvendes på in-situ fluorescenssensorer, som i øjeblikket er under udvikling, og vil derved kunne hjælpe os med at overvåge effekten af klimaforandringer på havets kulstofbudget endnu bedre i fremtiden.

LIST OF PAPERS

Paper A:

Anders D. Bruhn, Colin A. Stedmon, Jeromé Comte, Atsushi Matsuoka, Niek J. Speetjens, George Tanski, Jorien E. Vonk and Johanna Sjöstedt (2021): Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones. Front. Earth Sci. 9:640580. doi: 10.3389/feart.2021.640580

Published

Paper B:

Anders D. Bruhn, Urban Wünsch, Christopher L. Osburn, Jacob C. Rudolph and Colin A. Stedmon (2023): Lignin phenol quantification from machine learning-assisted liquid chromatography-absorbance spectroscopy data.

Submitted

Paper C:

Anders D. Bruhn, Urban Wünsch, Christopher L. Osburn, Rafael Gonçalves-Araujo, Mats Granskog and Colin A. Stedmon (2023): Dissolved lignin phenols across the Fram Strait: Towards in-situ measurements from fluorescence.

In last stages of preparation

TABLE OF CONTENT

ABBREVIATIONS

Ad/Al (S)	Ratio between acid and aldehyde for S phenols
Ad/Al (V)	Ratio between acid and aldehyde for V phenols
ACZ	Arctic coastal zone
AW	Atlantic waters
C	Cinnimalyl phenols
C/V	Ratio between C and V phenols
CDOM	Coloured dissolved organic matter
CO_2	Carbon dioxide
DAD	Diode array detection
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
EEM	Excitation-emission-matrix
FDOM	Fluorescent dissolved organic matter
GC	Gas chromatography
HPLC	High pressure liquid chromatography
LC	Liquid chromatography
mDOM	Marine dissolved organic matter
MS	Mass spectrometry
N-PLS	N-way Partial Least Square
P	p-Hydroxy phenols
PARAFAC	Parallel factor analysis
PARAFAC2	Parallel factor analysis 2
PCA	Principal component analysis
PLS	Partial Least Square
PPL	Polypropylene
PSW	Polar surface waters
P/V	Ratio between P and V phenols
S	Syringyl phenols
S/V	Ratio between S and V phenols
SPE	Solid phase extraction
TCC	Tucker Congruence Coefficient
TDLP	Total dissolved lignin phenol
tDOM	Terrestrial dissolved organic matter
V	Vannilyl phenols

BACKGROUND AND AIMS

BACKGROUND

Two unique and defining characteristics of the Arctic Ocean are the strong salinity stratification and the high concentration of terrestrial organic matter in surface waters. Both of these characteristics are caused by the geographical location of the Arctic Ocean. The Arctic Ocean is almost completely encircled by the landmasses of North America, Eurasia and Greenland. With the landmasses also comes a large river catchment. This leads to the Arctic Ocean receiving 11% of the global river discharge water, while only accounting for 1% of the global ocean volume (McClelland et al., 2012). The enormous input of fresh water also brings a large supply of tDOM and 14-24% of the DOC in the Arctic Ocean is therefore found to be terrestrial (Benner et al., 2005). In contrast, terrestrial DOC only represents 0.7% and 2.4% in Pacific Ocean and Atlantic Ocean, respectively (Opsahl & Benner, 1997). The supply of freshwater from the surrounding rivers is important for the formation of the halocline layer in the Arctic Ocean, a layer of water typically between 50-500 m depth with variable salinity and temperatures close to freezing point of seawater (Rudels et al., 1996). Without the halocline layer preventing warmer water masses to reach the surface, the perennial sea-ice in the central Arctic Ocean would not exist.

Another interesting aspect of the Arctic region is the increase in air temperatures with global warming. The Arctic region has warmed nearly four times faster than the rest of the globe over the last four decades (Rantanen et al., 2022). This large increase in air temperatures over such a short time span will have large consequences for the oceanographic conditions. With rising river discharge and increased erosion of permafrost soils, the supply of terrestrial carbon to the Arctic Ocean has already increased and will keep increasing in the future (Biskaborn et al., 2019; Woo et al., 2008; Nielsen et al. 2022). Permafrost holds ancient tDOM that can be highly bioavailable for microorganisms (Mann et al., 2015; Spencer et al., 2015; Vonk et al., 2013) and the increased flux of tDOM can therefore potentially impact carbon mineralization, the food web and the overall carbon cycling in the ocean.

The fate of tDOM in the Arctic Ocean therefore needs to be monitored more precisely in the future. Untargeted analysis, such as absorbance and fluorescence spectroscopy, on bulk seawater can be used to distinguish between terrestrial and marine sources (Stedmon & Nelson, 2015). The spectral fingerprint additionally carries information about mixing of water masses in the ocean and has already shown to be a useful tool to fractionate the contribution of freshwater coming from Arctic river catchments (Gonçalves-Araujo et al., 2016; Granskog et al., 2012; Stedmon et al., 2011). However, spectroscopy does not try to quantify how much tDOM the seawater contains. For this purpose, dissolved lignin can be used as a biomarker for tDOM. Lignin and lignin-derived compounds are considered to be major components of tDOM and is not naturally produced in seawater. Quantifying dissolved lignin has therefore shown to be excellent for tracing tDOM quantitatively (Hernes & Benner, 2003; Kaiser et al., 2017; Osburn et al., 2016).

The laboratory effort to measure dissolved lignin is however complex and time-consuming, involving steps like solid phase extaction, oxidation, chromatography and mass spectrometry. Predictive models utilizing absorbance and fluorescence spectroscopy have however shown to be able to provide an estimate of dissolved lignin in rivers and coastal waters (Fichot et al., 2016; Hernes et al., 2009). However limitations for these models often comes with lower concentrations of terrestrial DOC (Fichot et al., 2016) which is the case in open the ocean where lignin concentration are considerably low (Opsahl & Benner, 1997). The link between oceanic lignin concentrations and bulk measurements of absorbance and fluorescence therefore needs to be further investigated.

The fate Arctic tDOM has been studied in rivers (Holmes et al., 2008; Mann et al., 2015; Raymond & Spencer, 2015; Spencer et al., 2015; Vonk et al., 2013; Wild et al., 2019), but knowledge of its turnover in the extensive ACZ and in the open Arctic Ocean is more limited. In the light of accelerating global warming and increased erosion of permafrost soils in the Arctic region, some important questions about the fate of tDOM and terrestrial carbon in the still stands. Will the increased flux of terrestrial carbon into the ACZ lead to larger communities of marine bacteria, thereby impacting the foodweb and lead to increased production of CO_2 in the ACZ? Or will the terrestrial carbon persist, resisting degradation in the ACZ, and be exported from the Arctic Ocean to the Atlantic Ocean?

AIMS AND HYPOTHESES

To answer some of the above questions, this project studied the fate of ancient tDOM derived from permafrost soils in ACZs, to shed light on the degradation patterns taking place here before being transported further into the open ocean. Most studies on microbial degradation of permafrost soil has been performed with fluvial microbial communities and as bottle experiments. Bottle experiments creates an ever dynamic culture with change in both microbial and substrate compositions. Instead, one of the goals of this project was to investigate carbon turnover by marine microbes and their impact on the spectral properties, under steady-state conditions using a chemostat experiment. It was hypothesized that a part of tDOM provided by erosion permafrost soil is bioavailable to marine microbes and that different tDOM characteristics among glacial deposit types induce different growth rates and the establishment of different bacterial communities.

Next, the project aimed to develop a method to quantify tDOM in seawater. This was done in two stages. First the aim was to improve the applicability of HPLC combined with UV-visible absorbance to quantify individual lignin phenols after chemical oxidation of extracted oceanic DOM. The hypothesis was that both the sensitivity and selectivity of the approach could be greatly improved by incorporating a chemometric data analysis approach, such as parallel factor analysis 2 (PARAFAC2) which could circumnavigate current limitations with this particular chromatographic setup (HPLC-DAD) when analyzing specific compounds in complex DOM mixtures.

Secondly I investigated if the natural fluorescence properties of DOM in seawater could be used as a proxy for lignin content and character. More specifically if N-PLS of bulk fluorescence excitation-emission-matrix (EEM) measurements could predict dissolved lignin phenols as fluorescence of tDOM has a specific character. The goal here was to examine if this could pave the way for estimation of lignin phenol concentrations by in situ fluorescence spectroscopy with submerged probes.

1.1 WHAT IS DOM?

Dissolved organic matter is operationally defined as organic material that can pass a 0.2 µm pore size filter. DOM represents a large reservoir of organic material and as DOC in the ocean it exceeds the inventory of organic particles by 200 times, making it one of the largest reactive pools (Amon, 2004; Hansell et al., 2009). In the ocean, there is approximately 650-700 Pg carbon as DOC, which is comparable to the amount of carbon in our atmosphere, and therefore has to be considered an important component in the global carbon cycle (Fasham et al., 2001; Hansell et al., 2009).Despite its large inventory, DOC in the ocean exists at extremely low concentrations (34 – 80 µM), due to the large volume of the global ocean (Hansell et al., 2009).

Oceanic DOM is a very heterogeneous mixture and its magnitude and quality can vary greatly, influenced by a number of biological, chemical, and physical parameters (Koch & Dittmar, 2006). Oceanic DOM can be segregated based on its origins; marine and terrestrial DOM.

Marine DOM (mDOM) comprises of biomolecules released from living and decaying marine organisms which have also been altered chemically and biologically. The total production of DOM in the global ocean through primary production is about 60 Pg C yr^{-1} (Behrenfeld & Falkowski, 1997; Hansell et al., 2009).

Terrestrial DOM, is a mixture of vascular plant detritus, associated soil material, older fossil organic carbon (from carbonate rock erosion) and black carbon (e.g., soil organic charcoals and anthropogenic soot) (Hedges, 1992). Terrestrial DOM particularly differs from mDOM in the distribution of functional groups and aromaticity (Koch & Dittmar, 2006). Rivers supply 430 Tg terrestrial carbon yr^{-1} to the ocean (Schlünz & Schneider, 2000), which is less than 1% of marine DOC production. According to Romankevich et al. (2009) aeolian fluxes transport 96 Gg terrestrial carbon yr^{-1}.

1.2 CHARACTERISING DOM

Dissolved organic matter is thought to consist of hundreds of thousands of different molecules and untargeted measurements has shown that 1,500 to 7,500 of these can be assigned specific formulae in the range between 300 and 800 m/z, with the most abundant masses concentrated between 400 and 500 m/z (D'Andrilli et al., 2010; Koch et al., 2005). In general, although larger molecules are known to be present in DOM, multiple analyses confirm that most of DOM has a size smaller than 1 kDA (approximated to a molecular weight less than 1000 g mol^{-1}) (Benner & Amon, 2015; Hawkes et al., 2016). However even though these measurements reveal the elemental composition of DOM constituents, assigning the specific structure (not formulae) to all of these compounds is still not possible. Attempts to match the mass spectra with libraries only resulted in 1% of the compounds to be annotated (Petras et

al., 2017). Additionally, the analytical window limits the amount of compounds that can be analysed (see section 1.3.2) and combining multiple analytical techniques will be required to fully map the DOM pool in the ocean.

1.2.1 UV-Visible spectral characteristics of DOM

Both absorbance and fluorescence spectroscopy have showed great use in characterizing DOM. DOM can be categorized into fractions depending on its UV-visible properties (Figure 1), where coloured DOM (CDOM) is the fraction of the DOM that is coloured and fluorescent DOM (FDOM) is the fraction within CDOM that exhibits fluorescence. Studies have shown that the part of CDOM that fluoresces varies between 0.8% to 2.5%, depending on the wavelength of excitation (Andrew et al., 2013; Green & Blough, 1994). However not all compounds in the DOM and DOC pool are associated with material that absorbs and fluoresces (Figure 1).

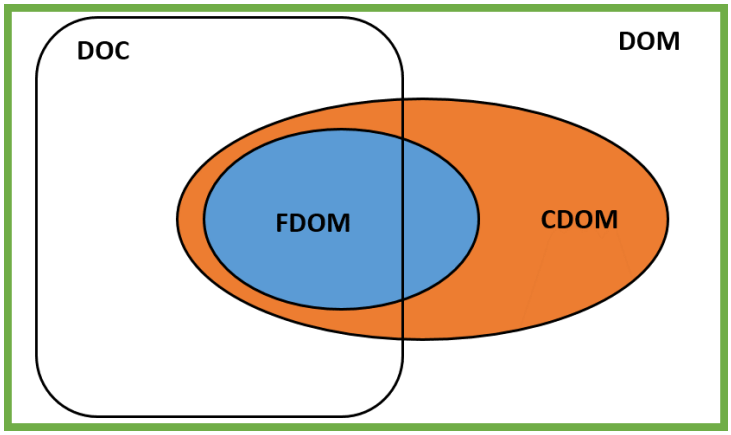


Figure 1: Fractionation of DOM depending on spectral properties. Not all DOM possesses spectral properties indicated by the white background. Orange circle indicates that some of the DOM is colored. The blue indicates that some of the CDOM fluoresces, namely FDOM. DOC is only a fraction of the entire DOM, but includes both colourless DOM, CDOM and FDOM compounds, indicated by black square.

For a specific chromophore, both the wavelength of absorption and the molar absorptivity increase with increasing conjugation or the presence of electron donors on aromatic rings (Stedmon & Nelson, 2015). However, the absorption spectrum of CDOM is often featureless and represents the combined absorption properties of the mixture (see Figure 2 for examples of absorption spectra). It can not provide specific information the composition of CDOM unless these signals are separated. The shape of the absorption spectra can however be characterized by fitting an exponential slope to the spectra. The magnitude of these slopes can be correlated to the molecular size distribution of the DOM and be used as a tool for differentiating between DOM sources (Helms et al., 2008; Stedmon & Markager, 2001).

The characteristics of FDOM can be mapped as an excitation-emission matrix (EEM) (Figure 2). Similar to CDOM, the emission and excitation properties of FDOM will be influenced by

increase in conjugation and aromaticity of their structure (Stedmon & Nelson, 2015). Fluorescence spectra of humic-like and terrestrially derived fractions typically display their fluorescence maximum at higher emission wavelengths (>400 nm) whereas the marine faction typical pose lower wavelength emission maximum (<400 nm) (Coble, 1996). EEMs can be characterised by comparing fluorescence ratios at different wavelengths (Coble, 1996, 2007) or by separating the combined signal into underlying independent signals using parallel factor analysis, PARAFAC (Murphy et al., 2013; Stedmon et al., 2003).

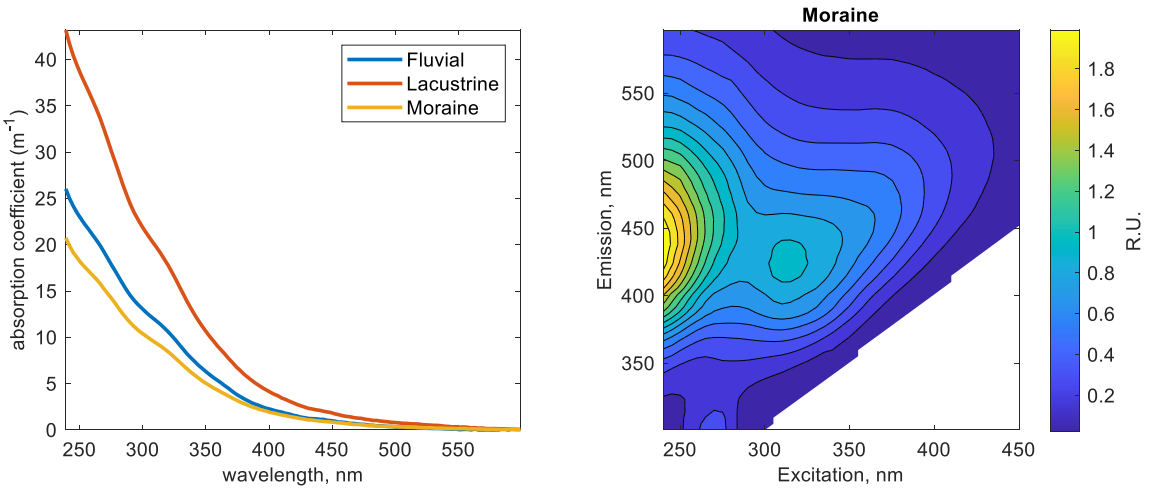


Figure 2: Absorbance spectra (left) and EEM (right) of respectively CDOM and FDOM derived from permafrost soils. The different soil types are fluvial (river deposits), lacustrine (lake deposits) and moraine (drained deposits).

1.2.2 Chromatography and detection techniques for DOM

The benefit of CDOM and FDOM measurements are that they are fast, economical and require low amount of sample. However, even though the analysis and interpretation of CDOM and FDOM spectra provides insight on contrasting composition of DOM in samples, it has limited specificity as the compounds or phenomena responsible are poorly resolved. Instead, measurement of specific biomarkers, although often involving more comprehensive laboratory procedures, can provide valuable additional measurements. These approaches often involve upconcentration of the DOM, followed by chromatographic separation, coupled with a detector, and finally comparison with known standards.

Physical separation of the upconcentrated DOM can be achieved by gas chromatography (GC) and liquid chromatography (LC). The separation can be based on different properties, such as polarity (Patriarca et al., 2018; Petras et al., 2017; Sandron et al., 2018), size of molecules (Hawkes et al., 2019; Wünsch et al., 2017), among others. In this thesis, the separation of DOM was performed based on polarity (the affinity of compounds to a non-polar material), which is also referred to as reverse-phase chromatography (Figure 3). For this approach, a polar mobile phase (often water) first transports the mixture of analytes (for example DOM) through a column consisting of non-polar material (stationary phase), where

all non-polar compounds are adsorbed. A non-polar mobile phase is then used to elute the compounds from the stationary phase. The gradual change between the two mobile phases causes the compounds absorbed onto the stationary phase to desorb according to their affinity towards the stationary phase and they will exit the column at different times (i.e. have different retention times on the column).

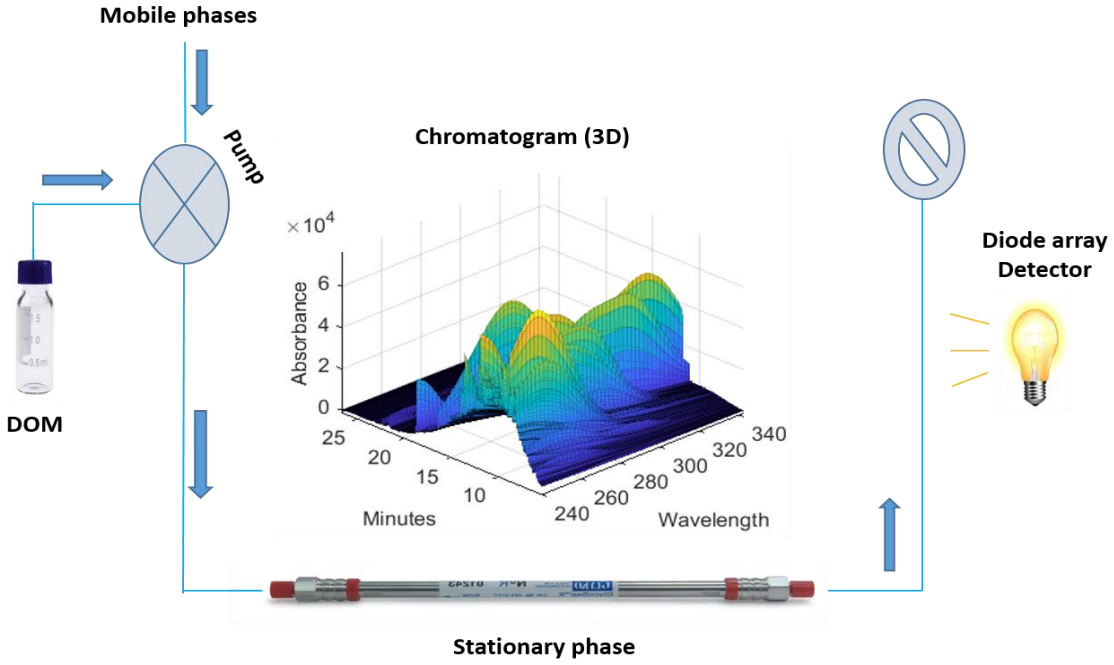


Figure 3: Example of liquid reverse-phase chromatography coupled with absorbance detection. The DOM sample is injected onto the stationary phase using a pump and a mobile phase. Changing the mobile phase from polar to gradually more non-polar leads to elution of specific compounds based on their polarity, thereby having different retention times. The output is a 3D chromatogram with retention time (minutes) and absorbance at different wavelengths (nm).

For detection of the separated compounds by LC, two different detection methods are most widely used, namely spectroscopy, in the form of absorbance and fluorescence detection, and mass spectrometry (MS). Spectroscopy relies on identification of specific compounds based on their absorbance or fluorescence properties, whereas MS determines the mass-to-charge ratio (m/z) of the eluting compounds or fragments. Both spectroscopy and MS provides spectra, which can be used for identification of compounds after comparing to standards.

In DOM science, LC coupled with spectroscopy has traditionally been used to investigate the overall polarity of a sample, by dividing the chromatograms into regions, and thereby identify the overall origin and biogeochemical processes of the extracted DOM (Caron et al., 1996; Mills et al., 1987; Parlanti et al., 2002; Koch et al., 2008). Liquid chromatography coupled with spectroscopy can also be used to fractionate the extracted DOM prior to other detection methods, such as mass spectrometry, and thereby create a hyphenated analysis between multiple analytical detection methods (Koch et al., 2008).

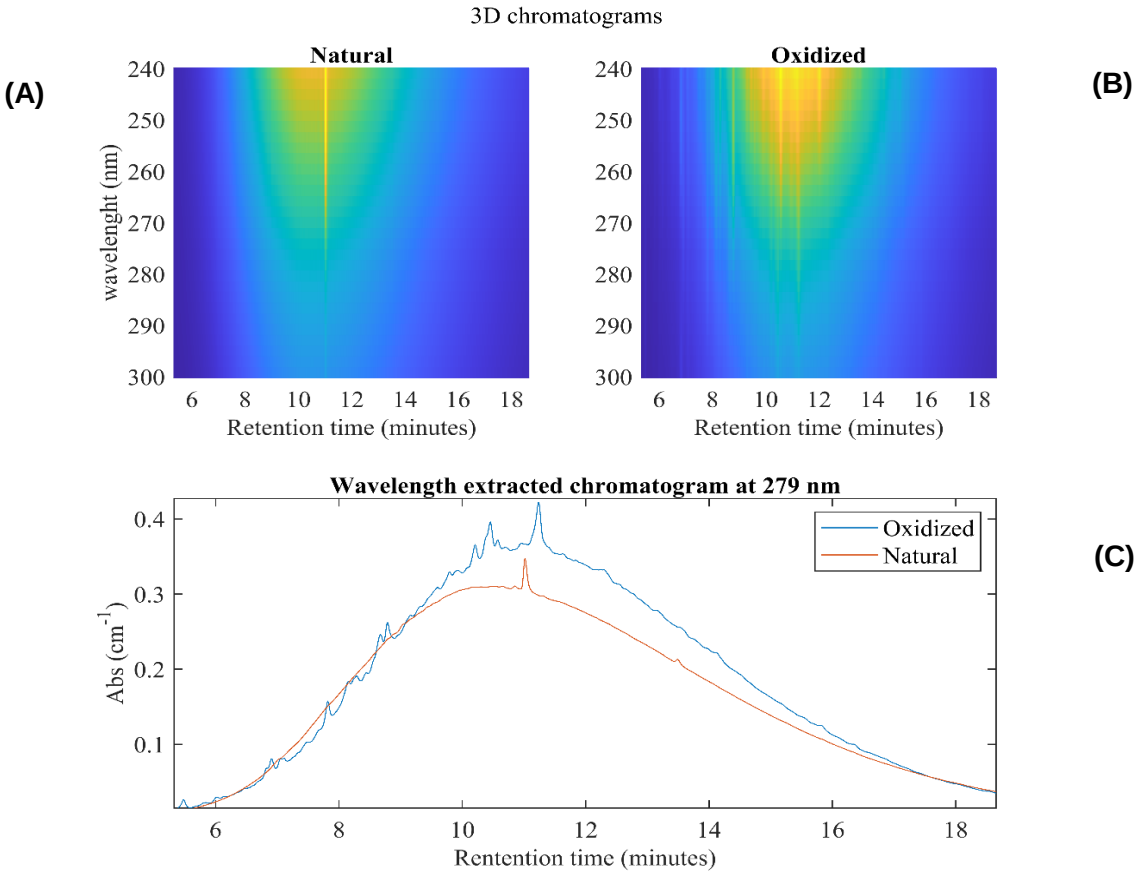


Figure 4: Absorbance chromatograms in 3D (A & B) and 2D (only 279 nm) for a natural and oxidized DOM sample measured on the HPLC-DAD. The DOM is extracted from tap water (ground water) as part of the method development phase.

As seen from Figure 4A&C the chromatogram of natural DOM does not contain many specific peaks and instead resemble the shape of a hill. This hill-like feature of DOM is due to its complexity of various compounds, that have similar chemical properties, functional groups, and molecular weights, which leads to spectral properties overlapping, which in turn makes it hard to identify specific spectra and compounds using absorbance. Oxidation of DOM can be performed prior to chromatography to cleave larger DOM compounds into their subunits, e.g lignin into lignin phenols (Hedges & Ertel, 1982; Lobbes et al., 1999; Kaiser & Benner, 2012; Yan & Kaiser, 2018). This process leads to an increase in the amount of chromatographic

peaks for oxidized DOM, compared to its natural DOM origin (see Figure 4B&C). Oxidation can therefore allow for a more detailed understanding of the underlining composition and properties of DOM, especially using chromatography coupled with spectroscopy.

Identification of specific molecules can represent a valuable set of biogeochemical tracers that can provide insights into the origins of the parent waters and the diagenetic changes that have occurred during transport, which has already proven to be important in terms of oceanographic (Goñi & Hedges, 1995; Hedges, 1992; Hedges et al., 1997; Benner, 2004; Benner & Opsahl, 2001; Hedges, 2002).

1.3 CHEMOMETRICS IN SPECTROSCOPY AND CHROMATOGRAPHY

The absorbance and fluorescence properties of DOM can be assumed to follow Lambert-Beer's law, which means that the overall intensity is equal to the sum of all of underlining spectra derived from a fixed number of individual compounds (Beer, 1852; Murphy et al., 2013).

As mentioned in section 1.2.2, chromatography can be used to physically separate the overlapping signals of individual compounds in DOM. Traditionally analytical chemists would focus on refinement of the chromatography, adjusting column properties, solvent mixtures, temperatures and pressure until the best separation is achieved. This works for simple mixtures but is a challenge for complex mixtures. For this, additional signal processing tools such as those in the field of chemometrics can be beneficial.

Chemometrics is a form of multivariate analysis useful for exploring and interpreting complex datasets, involving large numbers of variables that relate to one another in ways that are poorly understood prior to the analysis. Multivariate data analysis can be expanded to multiway data and is therefore well suited for chromatographic and spectral data. The central concept is that the combined signal can be reduced to a linear combination of underlying independent components (signals). In this thesis two approaches where tested and applied: Parallel Factor Analysis 2 (PARAFAC2) and n-way Partial Least Squares regression (N-PLS).

1.3.1 Parallel Factor Analysis

Before explaining PARAFAC2, it is essential to understand Parallel Factor Analysis (PARAFAC).

PARAFAC was formalized by Harshman (1970) and introduced to DOM biogeochemistry twenty years ago (Stedmon et al., 2003). PARAFAC is a multivariate data analysis technique that can decompose a complex three-dimensional dataset, such as EEMs (excitation x emission x sample) and chromatographic data (HPLC-DAD: retention time x wavelength x

sample), among others, into its underlying components and their scores (concentrations), making it easier to identify and quantify unique features (see Figure 5).

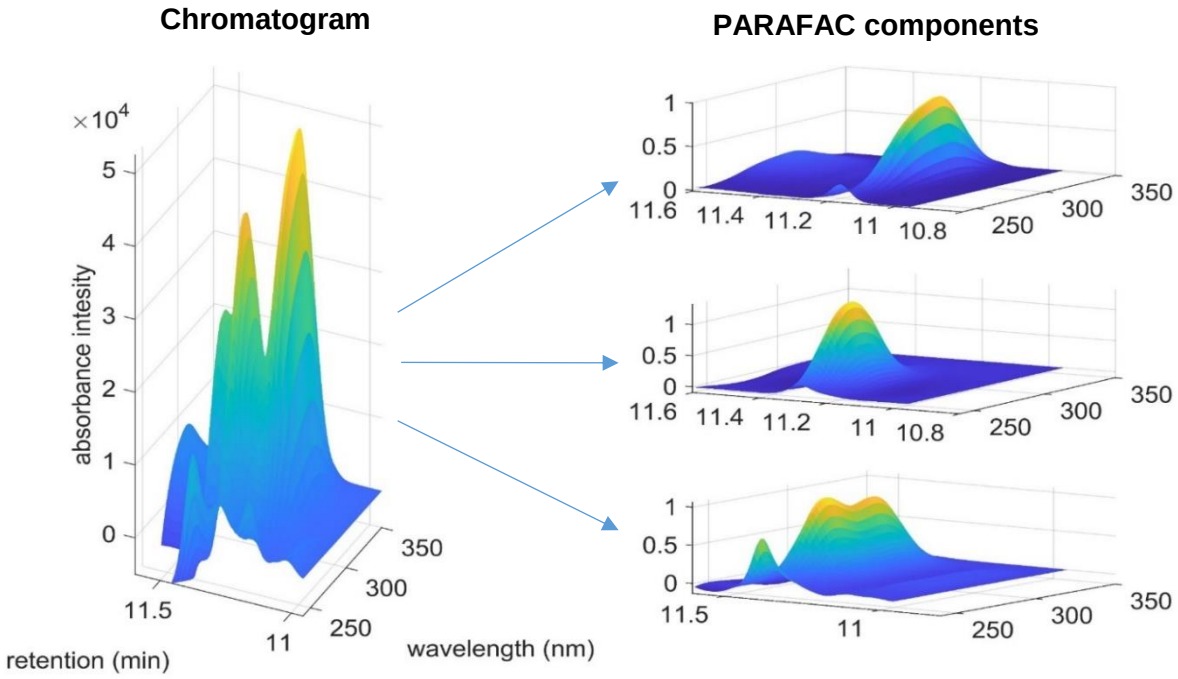


Figure 5: Conceptual example of PARAFAC decomposition of a chromatogram containing co-eluting peaks into three underlining components, each which reflects a unique compound as assessed by their absorbance spectrum.

In case of EEMs, each PARAFAC component consists of three loadings, an emission spectrum, an excitation spectrum and a concentration profile. The principles behind PARAFAC, follows some important assumptions about the nature of the data, such as variability, trilinearity and additivity (Murphy et al., 2013). In case of EEMs, the variability assumes that none of the fluorescent components can have identical spectra or covarying intensities. Trilinearity assumes that the chemical differences in each dimension of the dataset can be explained by the same number of underlying components. For EEMs, this for example means that the emission spectra are consistent regardless of the excitation wavelengths, while the excitation spectra remain constant regardless of the emission wavelengths, and that the fluorescence intensity increases linearly with concentration. The additivity principle assumes that the measured data is the sum of the contributions from each individual component. In the case of EEMs, this means that the fluorescence signal from each compound can be measured independently and added together to form the full EEM.

PARAFAC applies rotation constraints on components which ensures that the loadings and the concentration profiles are unique and interpretable. These constraints aim to reduce the ambiguity in the estimation of the components and facilitate the identification of the fluorescent

compounds. PARAFAC can additionally enforce non-negativity constraints on the components, which ensures that the concentration and spectral profiles are always positive, which is physically realistic.

The PARAFAC model can be described by the following equation:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} * b_{jf} * c_{kf} + e_{ijk} \qquad \text{eq(i)}$$

Where, x_{ijk} corresponds to the data points in a matrix, and in case of an EEM (excitation x emission x sample) *i* would correspond to excitation wavelengths, *j* corresponds to emission wavelengths, and *k* corresponds to sample. For an EEM, each PARAFAC component, *f,* is described by three vectors, *a*, *b* and *c* corresponding to, an excitation spectrum, an emission spectrum, and a concentration profile across samples. The residual (unexplained) signal is contained in e_{ijk}. Using an alternating least squares routine, the algorithm fits models until the improvement (reduction in e_{ijk}) between iterations falls below a given convergence criterion.

A modification of the PARAFAC algorithm, PARAFAC2, which was more suited to chromatographic data, was introduced by Bro et al. (1999) and Kiers et al. (1999). The application of PARAFAC2 has shown to solve problems with shifted, overlapping, and low intensity peaks in chromatography (Amigo et al., 2008; Skov & Bro, 2008) increasing the sensitivity and specificity of the detection.

For PARAFAC2 the following equation is applied:

$$x_{ijk} = \sum_{f=1}^{F} a_{if}^{k} * b_{jf} * c_{kf} + e_{ijk} \qquad \text{eq(ii)}$$

In case the HPLC-DAD, x_{ijk} corresponds to elements in the chromatographic data (retention time x wavelength x sample), where *i* corresponds to retention time, *j* corresponds to absorbance wavelength, and *k* corresponds to sample. As for PARAFAC, each component, *f,* is described by the three vectors, *a*, *b* and *c* corresponding to, the elution profile, absorbance spectrum, and a concentration profile across samples. However, in PARAFAC2, the superscript *k* in a^{k}_{if} allows the elution profiles in between samples to deviate slightly from each other to take into account minor remaining retention time shifts between samples. Similar to PARAFAC, the residual (unexplained) signal is contained in e_{ijk}. The alternating least squares routine is also used for PARAFAC2, where an algorithm fits models until the improvement (reduction in e_{ijk}) between iterations falls below a given convergence criterio.

An example of how PARAFAC2 modelling can complement and expand on the physical chromatography is shown in Figure 6. The top rows shows the chromatogram and spectra of three replicate samples (elution and spectral profile; see top of Figure 6). The bottom row shows the result of the PARAFAC2 analysis where the signal is split into four components, each with their own elution profile and absorption spectrum. Each sample has slightly shifted elution profiles, which shows how the algorithm is capable to compensate for small shifts in the chromatogram between samples. Additionally, a background component (yellow component in Figure 6) has also been isolated and removed from the calculation, which would otherwise interfere with the quantification of the targeted analytes.

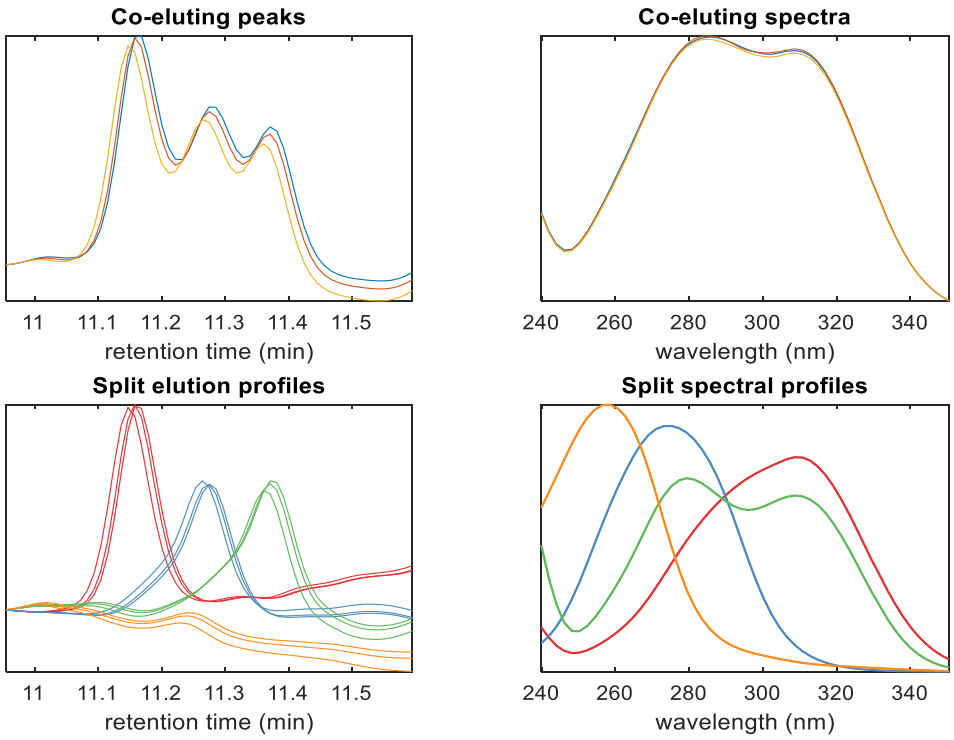


Figure 6: Interval from an absorbance chromatogram (HPLC-DAD) of co-eluting lignin phenols in three replicate injections of the same DOM sample. The absorbance of the spectra co-elute as well. Applying PARAFAC2 to the interval resulted in splitting of the co-eluting peaks and spectra into their individual profiles. Besides the three lignin phenols, a background component was isolated as well.

PARAFAC and PARAFAC2 has already been used to respectively to resolve EEM and HPLC-DAD data within multiple research fields, such as municipal water recycling, drinking and wastewater treatment, oil spill, food industry and medical sciences, besides natural DOM (Murphy et al., 2011; Yang et al., 2015; García et al., 2007; Vosough & Salemi, 2011; García et al., 2007). However, the use of PARAFAC2 to resolve lignin phenols measured on HPLC-DAD until now remain to be explored.

1.3.2 Partial Least Squares Regression

Partial least squares regression is an approach which can estimate a series of one or more dependent variables (Y) from a set of predictor variables (X) (Wold et al 2001). It is particularly suited to a situation where there are many predictor variables some of which can also be strongly correlated. It is preferential to multilinear regression, when dealing with spectral data, where signals from different (neighbouring) wavelengths are correlated.

First after appropriate pre-processing of the data, both matrices are decomposed independently into a set of scores and loadings (essentially two PCA models).

$$X=T \times P^T+E \qquad \text{eq(iii)}$$

$$Y=U \times Q^T+F \qquad \text{eq(iv)}$$

Where T and U are the scores for X and Y respectively, and P and Q are the respective loadings. The correlation between the scores for each can be calculated and expressed as a regression matrix (R).

$$U=T \times R \qquad \text{eq(v)}$$

This can then be combined with the loadings from the Y model to provide estimates of Y (Y') based on X scores.

$$Y'= U \times Q^T = T \times R \times Q^T \qquad \text{eq(vi)}$$

Maximum covariance between the U and T scores is found by running an iterative algorithm which focus on rotating the component (becoming latent variables) to minimise the difference between Y and Y' (Höskuldsson, 1988; Wold, 1975). When the covariance is maximized for one latent variable, the loadings and scores for this component are used to subtract data from X, and ocassionally Y, also known as deflation of the matrix, and a new PLS regression is performed on the residual X and Y data (Wold et al 2001).

When the predictor variables (X) are structured in a three-dimensional cube (i.e. EEM, excitation x emission x sample) rather than a two-dimensional matrix (as in PLS), N-PLS can be applied instead of PLS. N-PLS is designed to handle higher-order data and therefore maintain the dimensions of the data during decomposition, which allows N-PLS to extract more meaningful information from multiway data cubes, such as EEMs (Bro, 1996). Instead of decomposing X into a bilinear PCA model (as in PLS), N-PLS decomposes X into a set of

rank-one cubes, similar to the PARAFAC decomposition of EEMs. However, the correlation between the scores of X and Y follows the same procedure as for PLS (eq iii-vi).

The application of PLS and N-PLS has been widely used to predict concentration of specific compounds based on spectroscopy measurements (Bai et al., 2018; Hernes et al., 2009; Kumar & Mishra, 2012; Lin et al., 2022; Matero et al., 2010) which circumnavigates the use of more expensive and comprehensive measurements performed by more advanced analytical instruments.

1.4 LIGNIN: A BIOMARKER FOR TERRESTRIAL DOM

A great use of biomarkers in chemical oceanography is the ability to determine the origin of water masses and differentiate between them to understand distribution and circulation of these throughout the ocean basins.

Lignin is an amorphous, highly branched phenolic biopolymer unqiue to vascular plants, where it serves to strengthen the cell walls as a binder between cellulose and hemicellulose (Lewis & Yamamoto, 1990; Y. Lu et al., 2017; Monties & Fukushima, 2001). Lignin is the second most abundant biopolymer on earth (Norgren & Edlund, 2014) and the worldwide export of lignin from rivers to the ocean is estimated to be in the magnitude of 1.9 Tg yr^{-1} (Bao et al., 2015). The existence of dissolved lignin in the ocean, is an indication of its continual supply and slow degradation. Lignin has therefore shown to be has shown to be a powerful biomarker for the terrestrial material in aquatic systems, hereby in the ocean, and can be used to determine origin and diagenetic state of tDOM.

1.4.1 Lignin structure and chemistry

Lignin spans over a large interval in mass and size, depending on the source of plant material. However, the fundamental structure of lignin is the same regardless of the plant, source, or location, and it therefore provides the basis for lignin modification (Akpan, 2019).

Lignin arises from an enzyme-mediated polymerization of three phenylpropanoid monomers, coniferyl (1), sinapyl (2) and p-coumaryl (3) alcohols (see Figure 7; Dence & Lin, 1992; Katahira et al., 2018). The biosynthesis of lignin consists primarily of radical coupling (two unpaired electrons come together and make a bond) between the three aforementioned phenylpropanoid monomers with the addition of water or primary, secondary, and phenolic hydroxyl groups (Dence & Lin, 1992). The biosynthesis finally results in the formation of a three-dimensional polymer lacking the regular and repeated structure found in other natural polymers such as cellulose and proteins. As a result, lignin is a composite of chemically heterogeneous units.

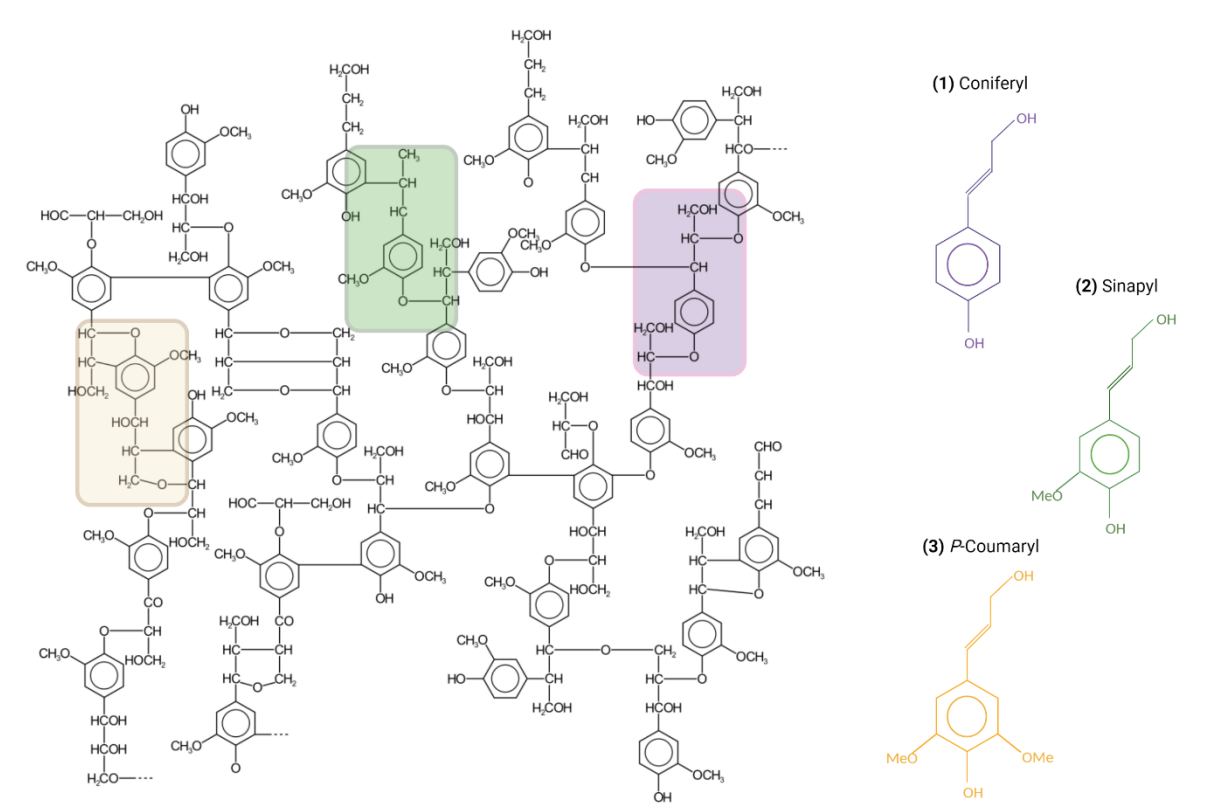


Figure 7: An example of a small section of the larger lignin structure, together with the three phenylpropanoid monomers which constitute the entire molecule in different configurations based on radical coupling. Created with BioRender.com.

There is no technique to quantify the entire lignin biopolymer and for measurement purposes within environmental sciences lignin is instead oxidized into phenol subunits by cupric oxidation. The cupric oxidation cleavages the C-O (ether) and C-C linkages between the monomer units in the lignin structure (Lu et al., 2016) and results in several phenolic monomers (known as lignin phenols) which can subsequently be quantified. These typically consist of aldehydes, ketones and acids of 4-hydroxyphenyl, vanillyl, syringyl, and cinnamyl phenols (see Figure 8).

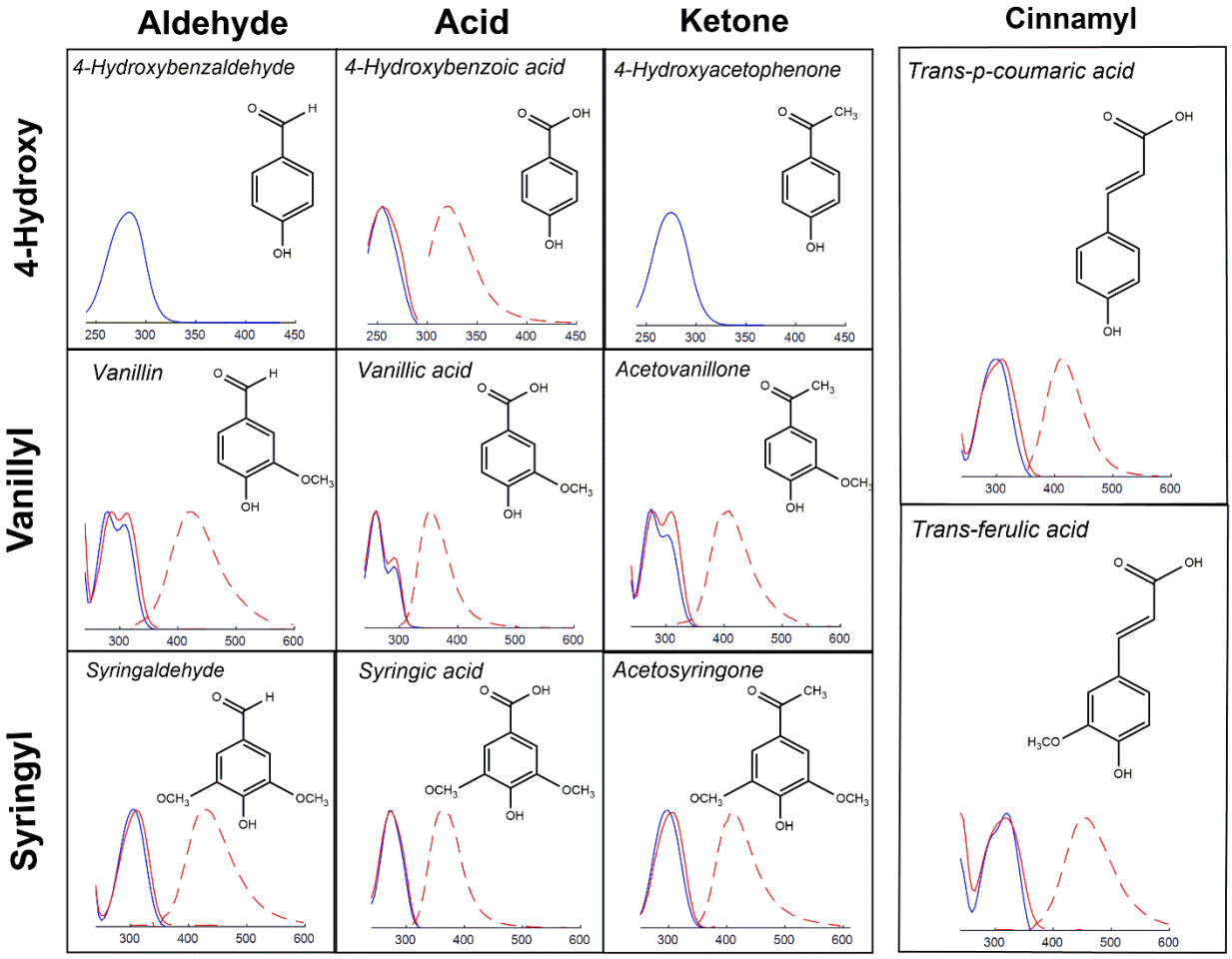


Figure 8: The eleven lignin phenols categorized by their chemical group (4-hydroxy, vanillyl, syringyl and cinnamyl) and carbonyl subunit (aldehyde, acid and ketone), together with their respective absorbance (blue), excitation (red solid) and emission (red dashed) spectra.

The ratio between lignin phenols can be used as an indicator for origin of terrestrial material (see Figure 9). The ratio of cinnamyl (C) to vanillyl (V) phenolic monomers (C/V) reveals the proportion of woody (trunk, roots, twigs) versus non-woody material (leaves, needles, grasses) with low ratio indicating higher proportion of non-woody (Hedges & Mann, 1979). The ratio of syringyl (S) to V, phenols (S/V) provides an indication of the contribution from angiosperm (flowering plants such as deciduous trees) versus gymnosperm (non-flowering plants such as conifers) material, with high ratio indicating more angiosperm origin (Hedges & Mann, 1979). S/V and C/V ratios can therefore be employed to link tDOM to its source material (Amon et al., 2012; Lobbes et al., 2000). The lignin phenol ratios therefore differs across plant species, however most significantly when comparing between families of plants (broad leaves, needle trees, bushes, flowers and so on) (Hedges & Mann, 1979). Lobbes et al. (2000) used this approach to characterize DOM in Siberian rivers and found that an increase in S/V and C/V ratios was linked to higher percentage of tundra within the drainage area.

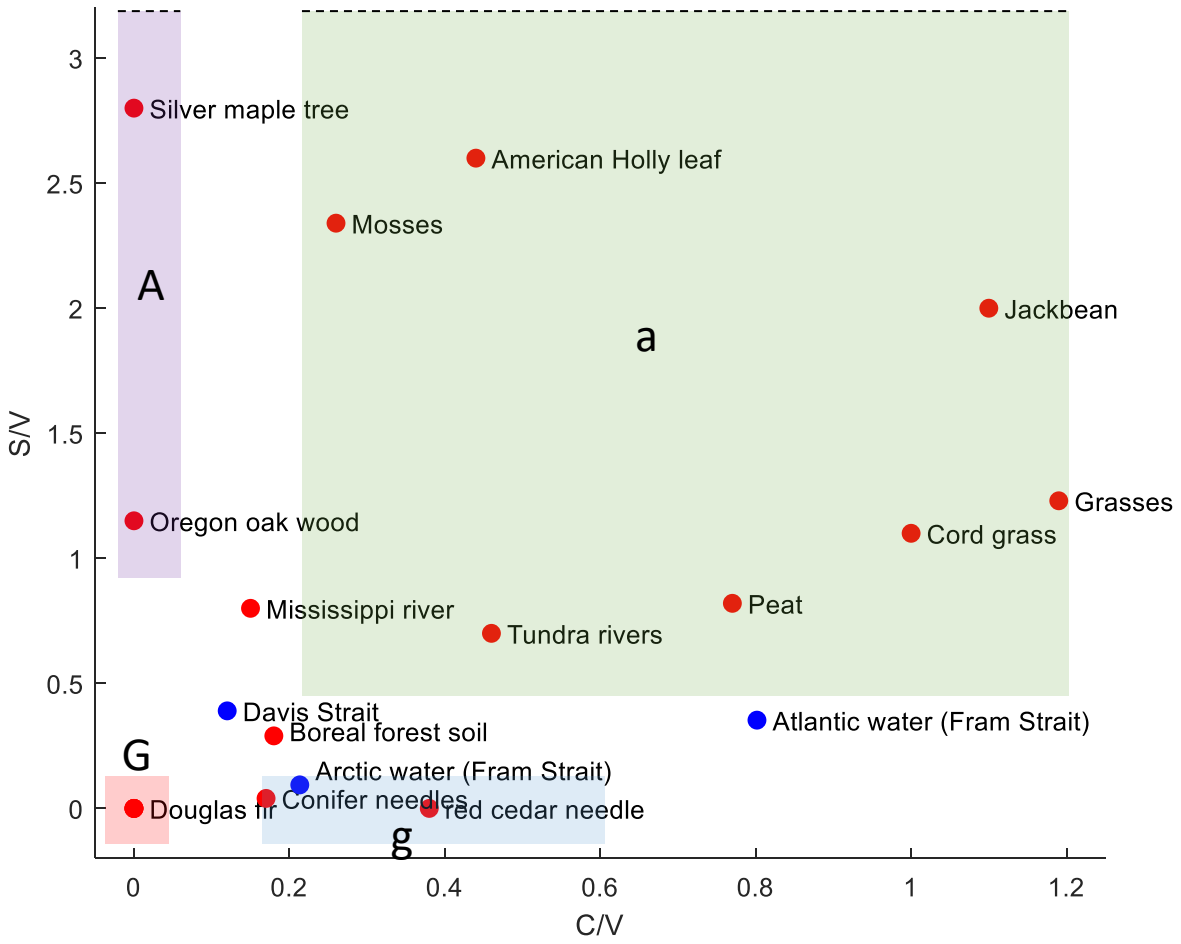


Figure 9: Origin of lignin in various plant materials and water masses, defined by the phenolic ratios between S/V and C/V (S=syringe,V=vanilyl and C=cinnamyl, see Figure 8 for phenolic monomers). The data for the red points belongs to other studies (Amon et al., 2012; Hedges & Mann, 1979) and are derived from plant tissues. The data for blue points is from Paper B and C included in the thesis. However, the blue dots are from ocean DOM. The colored boxes indicate the variation within different categories: (G) woody and (g) non-woody gymnosperm materials, whereas (A) woody and (a) non-woody material. The dashed lines in top of the blue and the green box indicates that the maximum limit for the S/V is higher than presented here.

1.4.2 Quantification of lignin phenols

The technique for characterization of lignin in environmental samples was first developed by Hedges & Ertel (1982) and focuses on the quantification of its composite lignin phenols, released during cleavage of the lignin macrostructure under alkaline cupric oxidation. In short, the method first involves upconcentration of DOM from environmental samples, thereby oxidation, under oxygen-free conditions and typically in an oven, followed by purification of the oxidation product to isolate the organic part, and then separation and quantification using chromatography coupled to a detector.

The technique has since its release been adopted and reworked by various research groups, and the most important improvements on the analysis are summarized in Spencer et al.

(2010). However, briefly they include reduction in laboratory steps and the introduction of new internal standards (Opsahl et al. 1999; Opsahl & Benner 1995; Hernes & Benner 2002), the use of an organic compound as antioxidant to encounter superoxidation (Louchouarn et al. 2000), change in oven temperature (Goni and Hedges 1992), option for solid phase extraction (SPE) instead of liquid-liquid extraction to purify organic part of the oxidation product (Kögel & Bochter, 1985; Yan & Kaiser, 2018b) and the replacement of toxic chemicals to more safer ones (Opsahl & Benner 1995). Finally, Yan & Kaiser (2018a) recently created a new method, where they modified the reaction vessel size, from 3.2 mL to 500 uL, which led to lower concentration of DOM and chemicals needed for the oxidation, and additionally changed the oxidant compound, from cupric oxide to cupric sulfate. Their new method showed to yield approximately 33% higher lignin phenol concentrations in environmental samples and at the same time increase the analytical precision.

Nowadays, the lignin phenols are mostly quantified with chromatography coupled with MS (Amon et al., 2012; Kaiser & Benner, 2012) or tandem MS (Reuter et al., 2017; Yan & Kaiser, 2018b). However, since all lignin phenols hold spectral properties, Lobbes et al. (1999) proved that HPLC-DAD indeed could be used to separate and identify lignin phenols in oxidized DOM samples. Detection by DAD offers an alternative and cheaper analysis to MS and tandem MS. However, HPLC-DAD is seldom used to measure lignin phenols because it lacks the specificity and sensitivity needed to compete with MS techniques, especially when dealing with a complex matrix such as DOM.

The concentration of total dissolved lignin phenols (TDLP) is often used as the indicator for the amount of dissolved lignin in a sample and is used for comparison between sample areas, DOM species and between studies. However, lignin also consists of other subunits, for example phenylcoumaran and stilbene structures, which can be released by acid treatment of lignin (Albinsson et al., 1999), but which are not quantified with the current techniques. TDLP therefore only used to address a part of lignin. The number of lignin phenols (out of the eleven ones) which can be quantified differs between lignin analysis techniques and therefore the TDLP abbreviation is often followed by a digit to account for this (for example TDLP11, is the sum of all eleven lignin phenols).

1.4.3 Lignin phenol predictions from optical measurements and chemometrics

Mutiple studies (Fichot et al., 2016; Mann et al., 2016; Hernes et al., 2006; Hernes et al., 2009; Hernes & Benner, 2003; Spencer et al., 2008; Spencer et al., 2009) have shown that it is possible to predict lignin phenol parameters from optical measurements. Hernes & Benner (2003) did the first attempt to use CDOM absorbance to predict TDLP by using the absorption at 350 nm across Mississipi River plume waters. Later did both Spencer et al. (2008) and Hernes et al. (2008) show that it was also possible for river waters as well. Hernes et al. (2009) showed for the first time, that also fluorescence could be used to predict the TDLP concentration alongside with the diagenetic state of lignin, more specifically S/V and C/V, by applying PLS modelling on unfolded EEMs (thereby breaking trilinearity). Hernes et al. (2009) found that PLS modelling on EEMs could help explain up to 91% of the lignin compositional and concentration variability in the samples. However, N-PLS modelling, which converse the

trilinearity of EEMs, still remains to be explored for predicting lignin phenols. Finally, Fichot et al. (2016) for the first time created an universal equation for predicting TDLP9 in coastal and shelf waters around the northern Hemisphere, for TDLP9 values ranging between 1-500 nM, using the natural logarithm for CDOM absorption at 250 nm. But at the same time, they found that their universal model had a hard time predicting TDLP9 values below 3 nM, which is often the case in the open ocean.

1.5 FATE OF TERRESTRIAL DOM IN THE OCEAN

Upon entering the coastal ocean, tDOM is exposed to a suite of processes, which either mineralise, transform or remove a portion of tDOM (see Figure 10). The largest processes are biodegradation, photodegradation and flocculation.

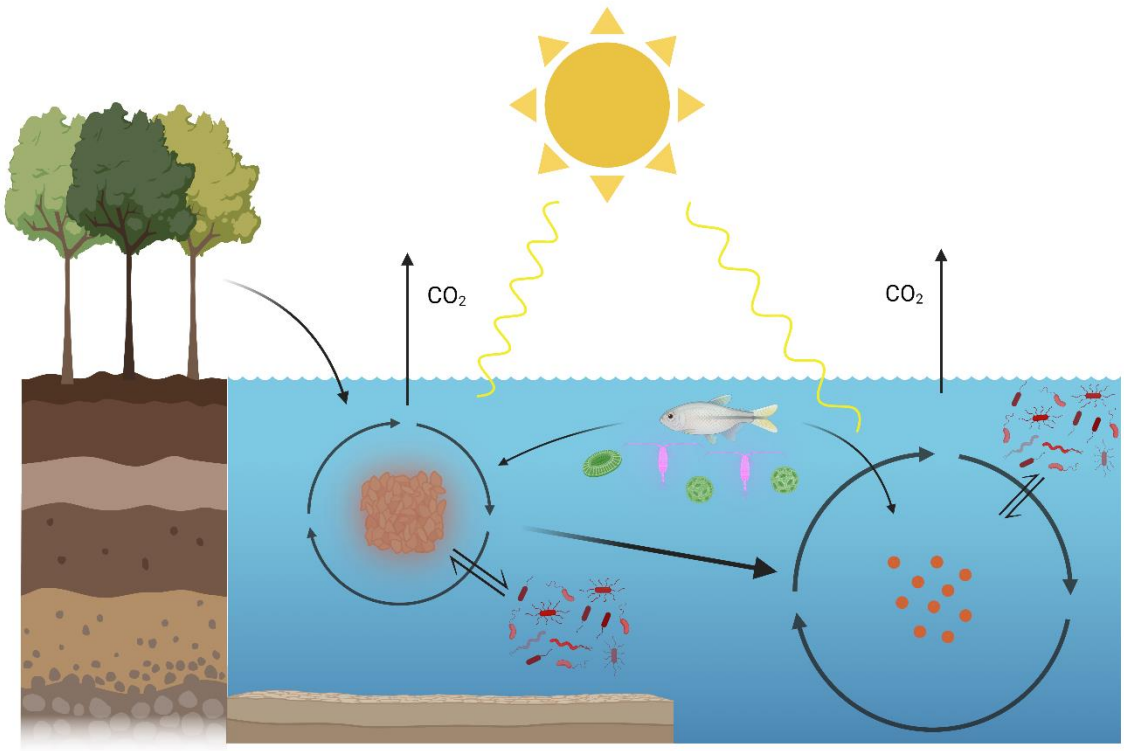


Figure 10: Overview of bio- and photodegradation of tDOM in the ocean. Both bio- and photodegradation can lead to direct mineralization of DOC to CO_2 or modify the tDOM into more labile parts, which in turn can be more biodegradable. The tDOM left after degradation in the coastal zone gets transported to the open ocean, where further degradation may occur. The turnover time is faster in coastal waters compared to open ocean, indicated by the size of the circles. Created with BioRender.com.

1.5.1 Biotic and abiotic processes

The reactivity of tDOM towards marine and photodegradation can be characterized based on time scales. Labile DOM is usually utilized a time scale less than a day, semi-labile DOM can last approximately 1.5 years, and DOM that survives more than a decade is considered refractory (Hansell, 2013).

Mineralization returns DOC to CO_2 and can be performed both microbally and photochemically (Cory et al., 2014; Cory et al., 2013; Müller et al., 2018; Tanski et al., 2019). Terrestrial DOM derived from permafrost soils has been found to be rapidly mineralized by marine bacteria (Müller et al., 2018; Tanski et al., 2019). Müller et al. (2018) found that 12% of the total DOC was mineralized within four days, and Tanski et al. (2019) found that the marine bacteria in their incubation experiments converted 2% of the total organic carbon delivered from permafrost soils into CO_2 per day. Interaction of tDOM with sunlight can also lead to complete mineralization, and Cory et al. (2014) found that photochemical processing of DOC in inland surface waters was responsible for one-third of the total CO_2 release. Additionally, Osburn et al. (2009) showed an overall photochemical mineralization of DOC at 20-30%, ranging from the Mackenzie River Delta across the shelf and into the Amundsen Gulf, with the river and estuary zone exhibiting higher rates of photodegradation than the shelf and gulf zone. However it is important to keep in mind that sunlight is limited by the turbidity and only appears in surface waters, whereas microbial degradation also occurs during the dark and therefore along the whole water column.

The interaction of tDOM with microbes and sunlight can also lead to partial degradation and transformation of tDOM (Cory et al., 2007, 2014; Grunert et al., 2021). Cory et al. (2007) found that the characteristic of DOM from Arctic tundra were related to residence time in water, hence exposure to aquatic microbes and sunlight. Grunert et al. (2021) demonstrated that tDOM along the continuum from an Arctic river to the ocean is altered by microbes and sunlight along the way, however with microbes most responsible for alteration in turbid river and plume waters, whereas photodegradation rapidly occurred in coastal waters turning. Studies have indicated that rates of photodegradation can exceed rates of microbial respiration for DOM in surface waters (Cory et al., 2013, 2014) and that photodegradation is more essential in influencing the chemistry of DOM in ocean surface waters (Cory et al., 2007). However, Grunert et al. (2021) found that photodegradation, while being a driver for transformation of DOC, was not responsible for net DOC remineralization. Degradation of tDOM in aquatic systems may therefore rather be viewed as a combination of two processes working on the back of each other.

When tDOM is microbially degraded the loss of tDOM may also go to biomass production. Sipler et al. (2017) showed that biodegradation of tDOM from Arctic rivers by marine bacteria can lead to 7-9% loss of DOC, including conversion to both biomass and CO_2, over the course of 6 days, while Mann et al. (2012) showed that as much as 30% of the DOC delivered from Arctic rivers is prone to biomass production and respiration during a month incubation. These studies, indicate that most of the DOC in tDOM is in fact semi-labile or refractory. The supply of tDOM can therefore lead to rapid changes in marine microbial community composition and growth (Bruhn et al., 2021; Blanchet et al., 2017; Sipler et al., 2017; Müller et al., 2018) and

tDOM can therefore be seen as an important component for the productivity in particularly coastal environments.

DOM can also be removed from its dissolved form due to flocculation (Sholkovitz, 1976; Sipler et al., 2017). Sholkovitz (1976) found that the proportion of riverine DOM that removed via salinity induced flocculation ranges between 3 and 11%, whereas Sipler et al. (2017) measured, specifically on extracted and filtered tDOM from permafrost soils, that 17% of the DOC was removed due to flocculation, when increasing salinity to 30%.

The part of tDOM that survives the processing in the coastal zone will be transported to the open ocean. Amon and Meon (2004) estimated that 44% and 65% of the annual Eurasian river DOC survives the coastal zone and will be exported to the Arctic Ocean. The efficient export of tDOM to the Arctic Ocean is also evident from the widespread distribution of this DOM in surface and halocline waters of the central Arctic Ocean (Amon et al., 2003; Stedmon et al., 2011) and in major Arctic gateways, such as the Fram Strait (Amon & Budéus, 2003; Gonçalves-Araujo et al., 2016; Granskog et al., 2012).

1.5.2 Experimental setup for biodegradability

Bacterial degradation of DOM can be studied by various experimental setups (batch, mesocosms, in-situ, chemostat etc.) Batch experiments are the most common approach for studying DOM biodegradation, particularly in Arctic aquatic systems (Blanchet et al., 2017; Müller et al., 2018; Paulsen et al., 2017; Sipler et al., 2017; Traving et al., 2017) and searching on Google scholar for “batch experiment bacteria DOM” gave 744 results, whereas “chemostat experiment bacteria DOM” resulted in “only” 66 findings.

The batch experiment simply consist of a closed bottle where the medium (containing substrate) and the microbial community is mixed together and let to incubate over a certain time span. Batch experiments can therefore be seen as a closed environment where the DOM pool change composition over time, which will likely lead to a constant change in bacterial community composition as well. Paulsen et al. (2017) specifically stated for their batch experiment to study long-term DOC biodegradability, that the results did not reflect in-situ conditions nor bacterial communities at the time of sampling, but rather was a quantitative measure of the overall biodegradability. Additionally batch experiments can vary in incubation time between studies and this often makes it hard to compare these (Marschner & Kalbitz, 2003).

An alternative to batch experiment, for studying biodegradation, is the chemostat experiment. The chemostat experiment was first introduced by Jacques Monod (1950), however got its name from Leo Szilard & Aaron Novick (1950). The chemostat experiment is a method for cultivating bacteria in a liquid medium, of fixed volume, where cells grow continuously under a constant medium replenishment, which is in contrast to the batch experiment where the medium is not exchanged. The central principle of the chemostat is that through the continuous addition of medium containing a substrate and simultaneous removal of culture, a stable equilibrium (steady state) where the substrate and the bacterial community is in

homeostasis is achieved (see Figure 11 for example). In this steady state, the rate at which the population of cells grows is equal to the rate at which the culture is diluted. An entire culture dilution once per day fits with the median growth rate of marine bacteria ranging from tropical to Arctic regions (Moriarty, 1986). The key advantages of the chemostat experiment is that it achieves a steady state with a constant defined environment and that the cell growth rate can be experimentally controlled by varying the rate of culture dilution. A chemostat environment can therefore me considered more natural-like to river and coastal environments where there often is a constant exchange in substrate and nutrients.

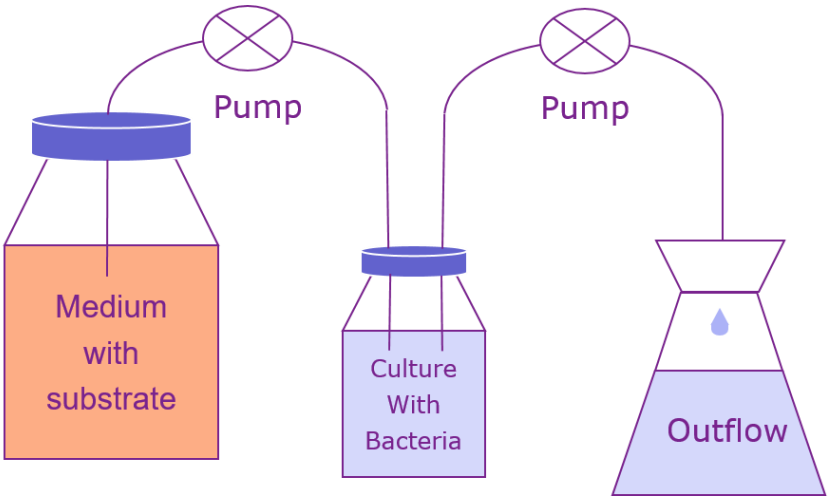


Figure 11: A classic chemostat setup, showing how medium is pumped into the bacterial culture, while at the same time culture is simultaneously pumped out and removed. The inflow and outflow runs at the same rate and therefore ensures steady-state conditions for the culture.

1.5.3 What happens to lignin phenols in the ocean?

The concentration of TDLP in the open ocean is generally below <1 nM (for high molecular weight lignin), while the coastal ocean ranges from 1 nM to 500 nM, depending on proximity to rivers and salinity (Opsahl & Benner 1997; Opsahl et al. 1999; Fichot et al., 2016). In the coastal zone, TDLP often have an inverse relationship to salinity and generally mixes linearly as it travels further away from rivers and estuaries and into the ocean (Hernes & Benner, 2003; Osburn et al., 2016). The lignin macrostructure contains a high portion of carbon (>60%; (Akpan, 2019)) and is therefore close related to the organic carbon content in terrestrial environments. The TDLP concentration have therefore also been found to be close related to DOC in riverine waters and a simultaneous decrease in lignin and DOC along estuaries has shown to be correlated (Hernes & Benner, 2003; Osburn et al., 2016). Lignin can therefore be used as a measure for distribution of freshwater and terrestrial DOC, as river water mixes with seawater and as a biomarker to study the flux of tDOM from land to oceans.

Transformation and degradation of dissolved lignin in aquatic systems is primarily driven by photodegradation, and can be assessed from the S/V ratio and the acid to aldehyde ratios (Ac/Ad) for V and S phenols (Hernes & Benner, 2003; Opsahl & Benner, 1998; Spencer et al., 2009). However, S/V as a parameter for photodegradation shows to be a complicated matter, as it has been found to both increase and decrease during exposure to sunlight (Benner & Opsahl, 2001; Hernes & Benner, 2003; Opsahl & Benner, 1998; Spencer et al., 2009). Ac/Ad (V) and Ac/Ad (S) ratios have showed to increase because of transformation of aldehydes to acids under photodegradation (Hernes & Benner, 2003; Spencer et al., 2009). However, this is not necessarily ubiquitous as Opsahl & Benner (1998) did not find any significant increase in Ac/Ad (S) after exposure of riverine DOM to solar irradiance for 28 days. Also, Opsahl & Benner 1998 found that increase in Ac/Ad (V) was exclusively for low molecular weight lignin and no difference was therefore observed for the high molecular weight part.

Lignin is thought to be very resistant towards microbial degradation and only a few fungi and bacterial species on land are known to be able to degrade it directly (Bugg et al., 2011; Dey et al., 1994; Hedges et al., 1988; Ramachandra et al., 1988; Zimmermann, 1990). However, degradation of dissolved lignin by microorganisms has been found to happen as riverine waters travel through the coastal zone, but this removal pathway was found to be at 3-5 lower rate than photodegradation and was additionally not found to alter the lignin phenol composition substantially (Hernes & Benner, 2003). Flocculation is also thought to be a large driver of lignin removal in low saline waters (Benner & Opsahl, 2001; Hernes & Benner, 2003). Benner & Opsahl (2001) estimated the loss of lignin phenols in low saline waters to be as high as 44% based on theoretical calculations. Losses of humic substances and tDOM at low salinities are generally known to be significant due to flocculation in river plumes (Sholkovitz, 1976; Sipler et al., 2017) and lignin should therefore still be able to reflect tDOM to a certain extent, even though flocculation occurs. In short summary, the composition of lignin phenols is not found to be altered by microbial and flocculation processes and the ratios are therefore primarily controlled by solar irradiance.

Once lignin has passed the coastal zone, no further diagenesis seems to occur as only small changes occur in C/V, S/V and Ac/Ad in the open ocean (Opsahl & Benner, 1997, 1998). This relates to the fact that the most susceptible part of lignin gets readily photodegraded in the coastal zone and only the less susceptible part reaches the open ocean (Hernes & Benner, 2003; Opsahl & Benner, 1998). The lignin structure in the open ocean is therefore considered more oxidized and structural different from its original version entering coastal waters, with a change from high molecular weight lignin to low molecular weight lignin with increasing salinity (Hernes & Benner, 2002, 2003; Opsahl & Benner, 1998).The persistence of dissolved lignin in the ocean therefore also varies depending on the distance from the coastal ocean, but its residence time is estimated to be on average 90 years in the open ocean (Hernes & Benner, 2002). The residence time is calculated using river discharge and ocean volumes, along with average lignin phenol concentrations (Opsahl & Benner, 1997).

1.6 ARCTIC OCEAN: CIRCULATION OF WATER MASSES AND DOM

Together with Greenland, Iceland, and the Labrador Seas, the Arctic Ocean is a key supplier of deep water in the Northern Hemisphere, and since this deep water contributes to global thermohaline circulation, it provides DOC to all deep waters in the world's oceans (Anderson & Amon, 2015). To address the sources, sinks, and distribution of DOC, it is necessary to examine water mass formation and circulation within the Arctic Ocean.

1.6.1 Water masses and circulation

The Arctic Ocean is a mix of saline water from the Atlantic Ocean and the Pacific Ocean, and freshwater from North American and Eurasian rivers together with sea-ice melt waters (see Figure 12 for map for overview of these water mass inputs).

Warm Atlantic water (> 4 °C) that pass through Fram Strait (left red line entering the Arctic Ocean in the bottom of Figure 12) will be cooled and meet the perennial sea-ice once it enters the Arctic Ocean. The upper 100 meters of these water masses will form lower salinity waters, namely lower halocline water, due to mixing with the meltwater from the perennial sea-ice (Rudels et al., 1996). The lower halocline water creates a barrier between the sea-ice and the warmer saline water underneath, which protects the sea-ice from melting, but also prevents deep-water formation in the central Arctic Ocean. The lower halocline runs along the shelf break towards the Laptev Sea, where it either turn north to return to the Fram Strait with the Transpolar Drift (large blue arrow running across the Arctic Ocean in Figure 12) or passes over the Lomonosov Ridge and flows towards the East Siberian Sea where it will meet waters with Pacific origin (Anderson & Amon, 2015).

Atlantic water masses that enter through the Barents Sea (right red line entering the Arctic Ocean in the bottom of Figure 12) will transfer their heat to the atmosphere and cool down. The increase in density results in most of these water masses running towards deeper waters through the St. Anna Trough (Schauer et al., 1997). However, some of the Atlantic water in the Barents Sea will keep running on the Siberian shelf and pass through the Kara Sea, further to the Laptev Sea and eventually all the way to the East Siberian Sea, before it will meet with the lower halocline waters and return to the Fram Strait with the Transpolar Drift (Jones et al., 1998).

The lower halocline waters and the Atlantic waters from Barents Sea (which pass through the Kara Sea, Laptev Sea and East Siberian Sea) will pick up freshwater and tDOM from major Eurasian rivers, as they run along the Siberian shelf. This will increase their DOC load, before returning together to the Atlantic Ocean with the Transpolar Drift and through the Fram Strait. A small part of these water masses may however export through the Canadian Arctic Archipelago and into the Davis Strait (Gonçalves-Araujo et al., 2016).

The intermediate Atlantic waters (yellow lines in Figure 12), below the lower halocline waters, ends up circulating the Arctic Ocean in cyclonic currents across most of the Arctic Ocean (Rudels, 2015). After circulation, these intermediate Atlantic waters will leave the Arctic Ocean through Fram Strait and are exported back to the Atlantic Ocean. Intermediate-depth water masses in the Arctic Ocean (in the depth range of 500-2000 m) largely follows the topography, resulting in several large loops (see yellow lines in Figure 12) within the central Arctic Ocean (Rudels et al., 1994).

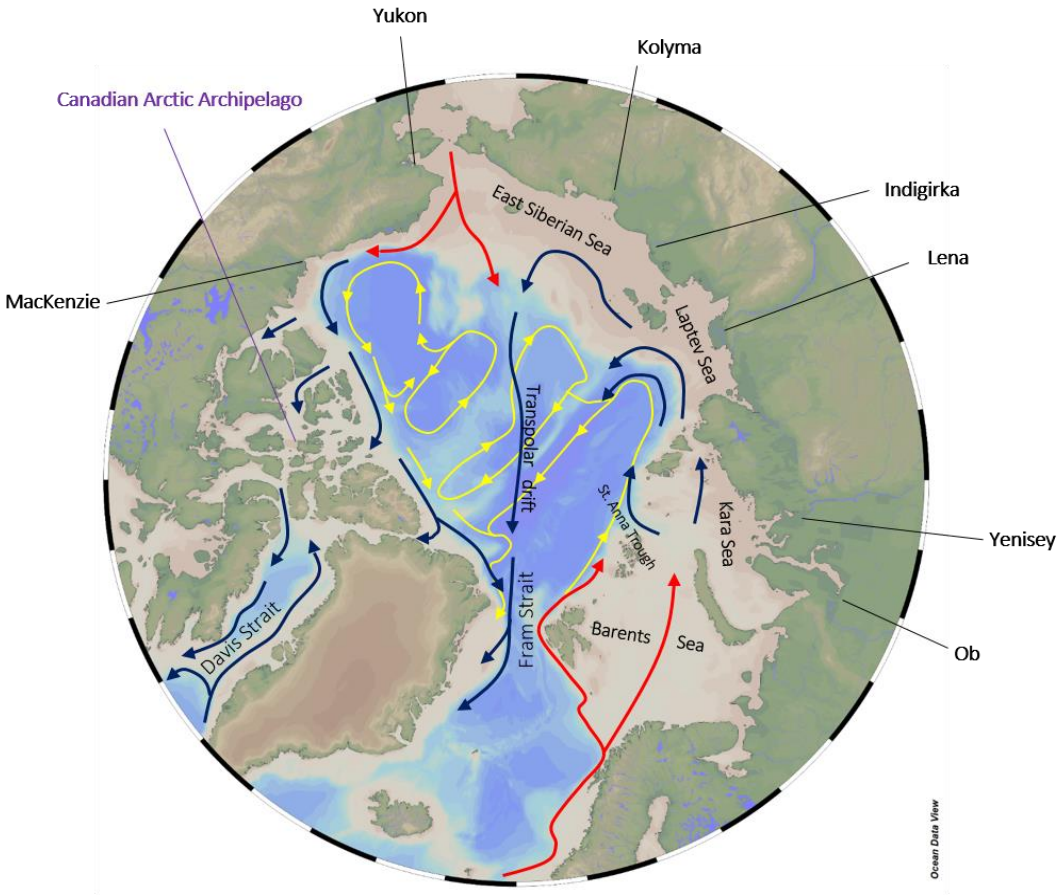


Figure 12: Orthographic map of the Arctic Ocean. The bold red and blue lines indicate warm and cold surface currents respectively. The yellow lines indicate deeper water currents and circulations. The black names indicate major rivers running into the Arctic. Map created with Ocean Data View.

Pacific water masses entering through the Bering Strait (red lines in top of Figure 12) are relatively fresh and therefore contribute mostly to the Arctic Ocean's upper layer (Anderson & Amon, 2015). However, a limited amount of high-salinity water created during sea-ice development in the Pacific sector of the Arctic Ocean penetrates to the deepest sections of the Canada Basin (Jones et al., 1995). Pacific waters that runs onto the East Siberian Sea (right red arrow in Figure 12) either connects with the Transpolar Drift and exits through Fram Strait or runs along the Canadian Shelf (left red arrow in Figure 12) and exports through both Canadian Arctic Archipelago and possibly through Fram Strait (Anderson & Amon, 2015).

1.6.2 Contrast between imported and exported DOM

Based on absorption coefficients and spectral slopes from CDOM measurements, three primary CDOM components in the Arctic Ocean has been determined: (1) a background marine CDOM pool; (2) a newly formed marine CDOM pool (3) the frequently dominant terrestrial CDOM pool (Stedmon et al., 2011). The Eurasian Basin is dominated by terrestrial CDOM while the Canadian Basin is more dominated by marine CDOM (Stedmon et al., 2011).

The DOC concentrations of Atlantic and Pacific inflow waters are 60 and 70 µM respectively (Amon et al., 2003; Anderson & Amon, 2015). As Polar surface waters (PSW) exports through the Fram Strait and the Canadian Arctic Archipelago, the DOC concentration increases to 80 µM and the DOM fingerprint of the water masses changes character (Anderson & Amon, 2015; Granskog et al., 2012).

The difference between DOM in Atlantic inflow waters and Arctic outflow waters, in the Fram Strait, is reflected in the substantially larger spectral slope in the 275-295 nm region of the absorbance spectrum for Atlantic waters, indicating photochemically changed material or a different CDOM source (Granskog et al., 2012). The slope ratio for the PSW is relatively low and equivalent to coastal waters reported by Helms et al. (2008), which emphasizes that the CDOM in PSW originates from more terrestrial sources than Atlantic waters. A substantial export of tDOM to the North Atlantic takes place through the Fram Strait and the Canadian Arctic Archipelago (Amon et al., 2003; Gonçalves-Araujo et al., 2016, 2020; Granskog et al., 2012; Walker et al., 2009) and transport of CDOM through the Fram Strait equals up to 50% of the total riverine CDOM input to Arctic Ocean (Granskog et al., 2012).

The reason for the change of carbon character for the Arctic outflow waters is that the surface water masses in the Arctic Ocean has picked up tDOM from continental runoff as they circulates around the Arctic shelfs. Ob, Yenisey, Lena, and Kolyma from Siberia, and Mackenzie and Yukon from North America, are the six major rivers in terms of discharge and watershed (Holmes et al., 2012). Arctic rivers deliver between 34 and 38 Tg yr^{-1} of DOC to the Arctic Ocean and surrounding basins in form of terrestrial source (Holmes et al., 2012; Manizza et al., 2009; Raymond et al., 2007). The Eurasian rivers export the largest proportion of tDOM to the upper Arctic Ocean as they together contribute about 84% of the annual DOC discharge to the Arctic Ocean, whereas the Lena river alone provide 36% of the total DOC per year (Amon et al., 2012).

The lignin flux is also highly related to discharge and the majority of lignin (49-78%) is released during the two months of spring freshet (Amon et al., 2012). Lena River has the highest TDLP concentrations all your around, however with freshet values exceeding 0.6 µM (Mann et al., 2016). In contrast, are the lowest TDLP concentrations found for Kolyma and Yukon, around 21 nM for both, during the winter (Amon et al., 2012).

Besides from being a recipient for tDOM, the Arctic Ocean also supports an average annual net primary production around 440 (± 21.5) Tg C yr^{-1} (Arrigo & van Dijken, 2011), but according to Mathis et al. (2007) only 10% of the carbon produced from net primary production accumulates as DOC (equal to approximately 44 (± 2.15) Tg DOC yr^{-1}). The total respiration

in the Arctic Ocean has been calculated to be around 51 Tg C yr^{-1} (Wheeler et al., 1997) and therefore exceeds that of the produced DOC from net primary production and can somehow indicate that a portion of the terrestrial DOC which comes from the river runoff may be consumed as well (Anderson & Amon, 2015).

1.6.3 Carbon cycling within the surface of Arctic Ocean

Adding the DOC from net primary production (44 (± 2.15) Tg DOC yr^{-1}) and the DOC from river runoff (36 (± 2) Tg DOC yr^{-1}; Holmes et al., (2012); Manizza et al., (2009); Raymond et (al., 2007)) together gives a input of 80 Tg DOC yr^{-} to the Arctic Ocean. If the total respiration of DOC to CO_2 by marine bacteria (51 Tg DOC yr^{-1} ; calculated based on Wheeler et al. (1997)) is subtracted from the summed DOC input, it leads to a surplus of 29 Tg DOC yr^{-1} in surface waters of the Arctic Ocean. The extra 29 Tg yr^{-1} of DOC, will potentially be exported to the North Atlantic Ocean through Davis and Fram Strait.

The DOC export in PSW through the Fram Strait has however been calculated to average 49 (± 6.2) Tg DOC yr^{-} over the last 12 years (Gonçalves-Araujo et al., 2020). The difference between the estimated surplus (29 Tg DOC yr^{-1}) and the export through Fram Strait (49 (± 6.2) Tg DOC yr^{-}) indicates that the uncertainties for many of these estimates may be relatively high. These numbers may be highly affected by seasons, as the conditions in the Arctic Ocean dramatically changes during a year. Finally, climate change has to be taken into perspective, as it will affect the carbon fluxes between reservoirs in the carbon cycle of the Arctic Ocean over time.

One particularly carbon flux in the carbon cycle of the Arctic Ocean that will be affected positively by climate change in the future, is the amount of DOC released by coastal erosion due to permafrost degradation (Nielsen et al., 2022). The annual flux of DOC from coastal erosion is estimated to be 55 Mg yr^{-1} (Tanski et al., 2016) and is therefore magnitudes lower than the flux from Arctic rivers (38 Tg yr^{-1}). However, the DOC in permafrost soils contain ancient carbon, which is immediately bioavailable (Mann et al., 2015; Vonk et al., 2013) and therefore can be rapid mineralized by marine bacterial communities once introduced into seawater (Müller et al., 2018). With accelerating global warming in the Arctic region, the magnitude of permafrost thaw will increase in the future (Günther et al., 2015; Nielsen et al., 2022) and so the amount of bioavailable DOC. The fate and the role of the terrestrial carbon in the ACZ and in the open Arctic Ocean is still uncertain in perspective to the microbial food web and the overall carbon cycle.

FINDINGS, DISCUSSION AND PERSPECTIVE

2.1 PERMAFROST DOM IN COASTAL WATERS AND CARBON CYCLING.

The thawing of permafrost along coastlines in the Arctic region can lead to the transportation of organic matter from permafrost soils into aquatic systems. The organic matter from coastal erosion can have a significant impact on the marine ecosystem, including changes in water chemistry, nutrient cycling, and the production and consumption of greenhouse gases such as methane and carbon dioxide by microbial activity (Schuur et al., 2015).

Before tDOM from coastal erosion reaches the shelves and the open parts of the Arctic Ocean it has to pass through the ACZ where it has the potential to be removed or transformed. Coastal permafrost stores large amounts of carbon, which can potentially be readily mineralized by microbes (Tanski et al., 2017, 2019). Changes in supply of carbon to ACZ can lead to changes in the abundance and diversity of marine microbial communities, which can have cascading effects on the marine ecosystem by influencing the competition for nutrients between heterotrophic bacteria phytoplankton (Thingstad et al., 2008). Multiple studies have already demonstrated that the addition of tDOM from Arctic rivers can lead to significant changes in the composition and growth of marine microbial communities (Blanchet et al., 2017; Sipler et al., 2017; Müller et al., 2018). Despite these findings, there are still gaps in our understanding of the impact of tDOM from coastal erosion on marine microbial community composition and carbon processing.

Paper A therefore investigates the interaction between tDOM and marine microbes, using three different permafrost soil types, also referred to as glacial deposit types. To test carbon processing and changes in marine microbial communities, a chemostat approach was chosen for this study, rather than batch experiments. This was done in order to replicate local substrate supply conditions. In batch experiments, the microbes are grown in a closed environment where the nutrients and waste products are not continuously removed or replaced. As the microbes consume the available nutrients and produce waste products, the composition of the substrate changes, which can affect the growth and metabolism of the microbes and thereby lead to changes in the population size and overall activity.

The results from the chemostat experiments in Paper A provide evidence of how differences in the tDOM quality between three permafrost soil types leads to the development of distinctive marine microbial communities. The difference in tDOM quality also leads to different biodegradability, with much higher bacterial growth efficiency for moraine deposit types (66%) compared to fluvial and lacustrine deposit types (13 and 28% respectively). This indicates that the tDOM quality of eroding permafrost soil types along Arctic coastlines ultimately controls how much of the terrestrial DOC which will be stored as biomass and how much will be respired as CO_2 in ACZ. Such difference in the carbon processing of different permafrost soil types, is currently not included in calculations of DOC fluxes from coastal erosion to the ocean, but should be considered for future modelling. However, the carbon processing of the

terrestrial DOC by marine microbes is generally not included in the land-to-ocean carbon flux calculations (Nielsen et al., 2022; Terhaar et al., 2021).

The findings in Paper A demonstrate that an increased addition of tDOM to ACZ will lead to an overall increased abundancy of heterotrophic microbes. The increase in heterotrophic microbes will potentially alter the structure of the food web, more specifically the balance between autotrophs (phytoplankton) and heterotrophs (bacteria), by increasing the competition for inorganic nutrients (Thingstad et al., 2008). However, other studies have shown that phytoplankton can later benefit from regenerated nutrients released from microbes processing tDOM (Sipler et al., 2017; Traving et al., 2017). Tank et al. (2012) also found that mineralisation of riverine dissolved organic nitrogen supports nearshore primary production. Other studies have shown that increased supply of tDOM into coastal areas will cause an overall reduction in coastal phytoplankton production due to light absorption by DOM (Balch et al., 2012; Wikner & Andersson, 2012). The question on how the addition of tDOM from permafrost will ultimately influence the carbon cycling and food web is therefore complex, but it is clear that both the quality and quanity of tDOM play a key role in shaping light conditions, carbon mineralisation rates, and nutrient availability.

The high degree of reproducibility between the replicate chemostat experiments, provides confidence in the robustness of the findings and the experimental approach. The entire culture was exchanged once per day, which fits with the median growth rate of marine bacteria (Moriarty, 1986). It can however be discussed if this exchange rate fits with local conditions, which likely vary a lot depending on season and location, and at the same time does not account for episodic coastal erosion events. The exclusion of additional substrate, such as marine DOM may also have implications for the microbial community composition and the overall carbon processing. Sipler et al. (2017) for example showed a lower rate of marine microbial decomposition of tDOM from permafrost soil (7-8% of terrestrial DOC removed), which they, among other factors, attribute to the fact that they included sunlight and phytoplankton growth in their experiments. However, the DOM produced by phytoplankton may have masked the removal of tDOM they observed. Further chemostat experiments, including marine DOM, phytoplankton and light should therefore be carried out in the future to undercover some of these concerns.

2.2 SOLID PHASE EXTRACTION OF DOM IN SEAWATER

Characterisation of DOM in seawater often requires concentration of material through solid SPE. Isolating a representative portion of the original DOM is often impossible as the extraction is selective. First, the acidification of the sampled water in order to increase extraction efficiency of carboxylic- and phenolic-rich compunds, can denature proteins and peptides and change the reactivity for some classes of compounds. Next, the selectivity of compounds during SPE depends highly on the sorbent used (Arellano et al., 2018; Dittmar et al., 2008; Perminova et al., 2014). Polypropylene (PPL) cartridges are often selected as the sorbent of choice resulting in high extraction efficiencies of DOC (Arellano et al., 2018) and acceptably representative DOM composition (Perminova et al., 2014), as it is capable of

extracting molecules with a range of polarities (Dittmar et al., 2008). PPL cartridges was therefore chosen for lignin isolation from seawater for both Paper B & C.

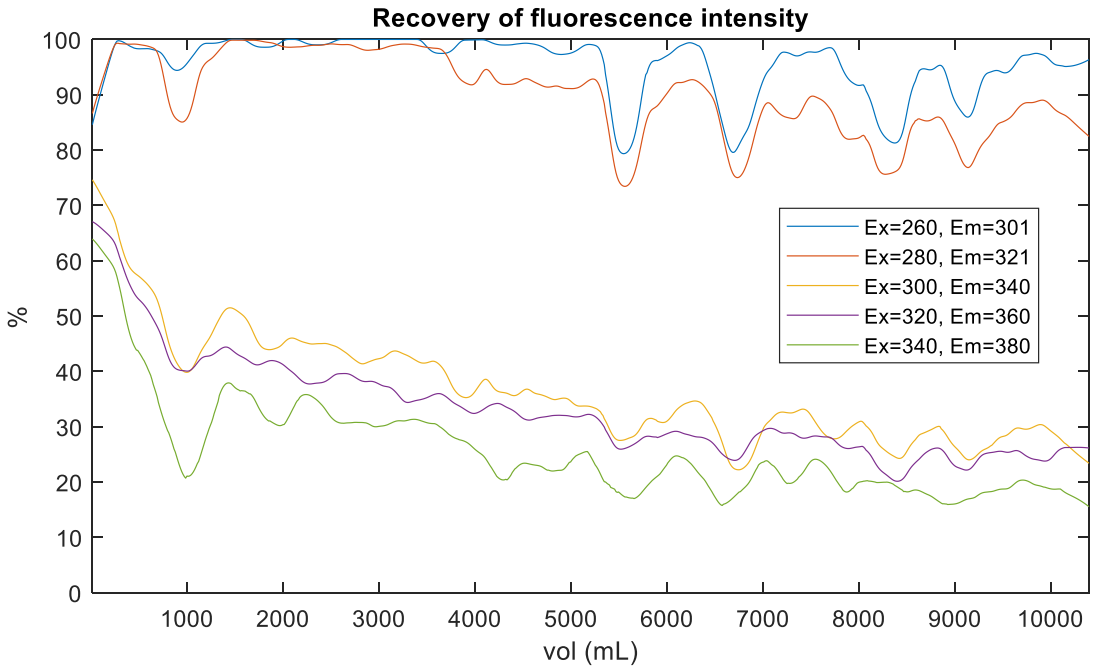


Figure 13: Recoveries of fluorescence intensity at different emission and excitation pairs during SPE of seawater from the Davis Strait through a PPL cartridge. The permeate seawater from the SPE was connected to a flow cell in a fluorometer (AquaLog, Horiba) and EEMs were measured every third minute. The recovery is calculated as the difference in fluorescence intensity from the fluorescence of the seawater prior to extraction.

To test the extraction efficiency of DOM in seawater and learn more about changes in the quality of the extracted DOM, 10 L of seawater collected in the Davis Strait (same water mass used for Paper B) was passed through a PPL cartridge at a steady flowrate (4 mL/min). The PPL cartridge was connected online to a sealed quartz cuvette inside a fluorometer (AquaLog, Horiba) and fluorescence EEMs were continuously measured on the permeate water every third minute. To speed up the fluorescence measurement, emission spectra (210-620 nm) were only measured at five excitation wavelengths (240, 260, 280, 300 and 320 nm).

The extraction efficiency for UVA fluorescence (Ex$\leq$280 nm and Em$\leq$321 nm) was above 80% for the first 5 L (Figure 13). For visible fluorescence (Ex$\geq$300 nm and Em$\geq$340) the extraction efficiency steeply dropped towards 1 L (Figure 13) then only slowly decreased with increasing volume. The extraction efficiency did not show a decline towards 0 % during the whole extraction, indicating no saturation of the sorbent material was reached with the use of 10 L of seawater.

To investigate a change in character of DOM during the extraction of 10 L seawater, using PPL cartridges, normalized emission spectra at excitation 320 nm for each time point (n=770) were compared to each other (Figure 14). As seen from Figure 14A the normalized spectra look very similar across the whole extraction. However the position of peak max for the

emission spectrum (at excitation 320 nm) changes “dramatically” after the extraction of 60 mL, but afterwards only slowly changes for the rest of the extraction (see Figure 14B). Even though peak max kept changing, the spectral similarity (defined by the Tucker congruence coefficient (TCC); Lorenzo-Seva & ten Berge (2006) and Tucker (1951)) between the emission spectrum (at excitation 320 nm) from the extracted seawater and the original seawater remained identical to each other posing a TCC value equal to 1 during the whole extraction (see Figure 14B).

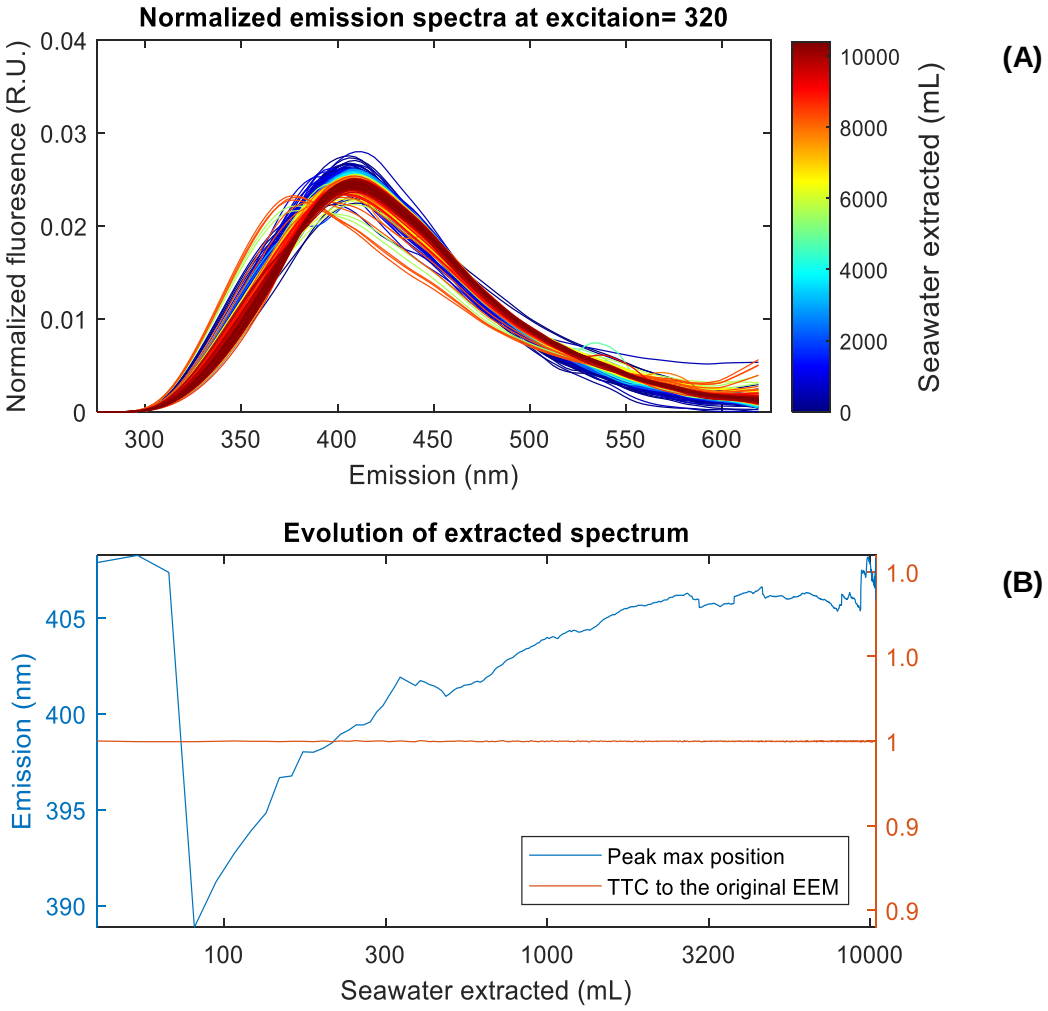


Figure 14: (A) Normalized emission spectra (normalized to area under the curve) taken at excitation 320 nm for the permeate water during the SPE of a seawater sample from Davis Strait (10 L) using a PPL cartridge and (B) the peak max wavelength for the emission spectrum and TCC value to the spectrum prior to extraction across the whole extraction. Note that the x-axis in the (B) is log-scaled.

The assessment of the DOM quality and recovery during SPE shows that no break-through seems to be achieved during the extraction of 10 L of seawater, as the extraction efficiency did not indicate a steep decline, while the spectral character of the extracted water at the same

time remained identical to the original seawater. The slight change in peak max around 60 mL, could however indicate two phases during the SPE. The recovery for the visible fluorescence fits with the findings by found other studies (Dittmar et al., 2008; Wünsch et al., 2018) whereas the recovery for the UVA fluorescence seems to be higher than previously reported by Wünsch et al. (2018).

2.3 IMPROVING LIGNIN QUANTIFICATION BY HPLC-DAD

To try to improve the HPLC-DAD method proposed by Lobbes et al. (1999), a new cupric oxidation technique from Yan & Kaiser (2018a) was adopted and slightly modified. This new cupric oxidation technique makes use of $CuSO_4$ (instead of CuO) as oxidant and uses a smaller reaction vessel volume of 500 uL (instead of 3.2 mL). The new oxidation method has improved precision of quantification and yield of lignin phenols compared to earlier cupric oxidation techniques. It has additionally led to the need of lesser sample volumes needed for the extraction of DOM using PPL cartridges (Yan & Kaiser, 2018a). To create a cupric oxidation method around the one proposed by Yan & Kaiser (2018a), stainless steel reaction vessels (850 uL) were specially designed and produced (Figure 15). The size of the reaction vessel differed from Yan & Kaiser (2018a), but volumes of chemicals added to the reaction vessel were adjusted to achieve similar concentrations.

Figure 15: Stainless steel reaction vessel with a volume 850 uL.

As seen in Figure 4 and in Figure 16A, the chromatograms of cupric oxidized DOM from HPLC-DAD suffer from elution of a broad background component. As part of the method development in Paper B this interference was removed by transforming the chromatogram into its 2nd derivative (see Figure 16B). This also has the added benefit of aiding peak detection algorithms subsequently applied to decompose the chromatogram (Vivó-Truyols et al., 2005).

As seen from Figure 16B, the 2nd derivative chromatograms contains high degree of co-elution between peaks. In Paper B, the lignin quantification method was therefore further improved by automating the process of resolving the co-elution by applying PARAFAC2, which

decompose the overlapping peaks into unique components. The approach improves the sensitivity and specificity of most lignin phenols, compared to manual integration of peaks, by removing interfering signal from neighboring peaks and any residual background DOM. The results in Paper B therefore shows that lignin quantification with HPLC-DAD can be slightly improved using PARAFAC2 and detection limits below 2 nM for most lignin phenols can be achieved, which is similar to newer MS techniques (Reuter et al., 2017; Yan & Kaiser, 2018b). The interlaboratory comparison on seven samples from the Florida Everglades, where GC-MS measurements were performed at North Carolina State University, showed that the results obtained with the HPLC-DAD using our 2nd derivative/PARAFAC2 method for analysis provide somehow similar results. However, HPLC-DAD tended to slightly underestimate most the individual lignin phenol concentrations. The concentrations of the summed V and S phenols and TDLP11 actually showed to fit better. Anyhow, the benefits of HPLC-DAD compared to GC-MS is that HPLC-DAD is cost-effective and eliminates the need for derivatization of the DOM prior to analysis. The method developed in Paper B can therefore provide scientists and laboratories with an option to measure lignin in the ocean competitive to MS techniques.

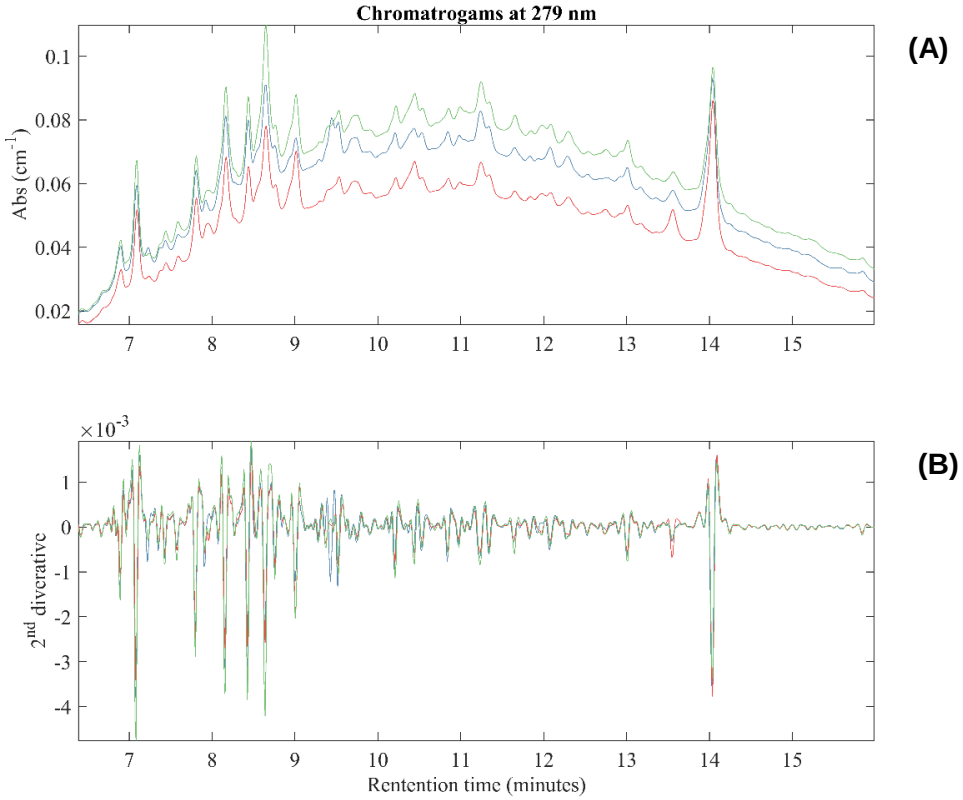


Figure 16: Three wavelength extracted chromatograms at 279 nm absorbance (A) and the 2nd derivative transformation of these (B) for three cupric oxidized DOM samples (Fram Strait)

As seen from Figure 16B, the oxidized DOM chromatogram is full of prominent peaks, besides the ones belonging to the lignin phenols. In fact, PARAFAC2 found additional spectra in the intervals of the chromatogram where the lignin phenols were eluting. These spectra often resembled those of single-analyte compounds rather than residual DOM background or noise. This implies that the cupric oxidation of DOM provides the release of additional compounds and reaction products, from oxidative cleaving of larger compounds, which potentially can be used as biomarkers to describe DOM in the ocean. It would therefore be interesting to apply the machine learning-assisted method to more intervals in chromatogram and maybe uncover these biomarkers. Baccolo et al. (2021) has recently automated the application of the PARADISe (a software utilizing PARAFAC2) for untarget GC-MS analysis, so that it can be applied over the entirety of the chromatogram and extract all relevant spectra, elution profiles and concentration of relevant compounds. It would be interesting to apply the automated version of PARADISe to the chromatograms obtained by HPLC-DAD in this thesis and see if the same concentration of lignin phenols could be quantified to validate the developed 2nd derivative/PARAFAC2 method and additionally investigate the potential for new biomarkers.

The newly developed lignin method in Paper B shows that we can reduce the volume of seawater needed for SPE, down to 1 L and probably less. Sampling of lignin in the ocean has traditionally been a very heavy task and required huge volumes of seawater, since concentrations often are low, and earlier studies have sampled as much as 5-30 L (Kaiser et al., 2017; Opsahl et al., 1999). These volumes often limits the amount of samples taken for lignin analysis and therefore affects the spatial resolution. The water budget on research vessels is often tight and the reduction in volume will therefore also serve to be more economical. Additionally, experiments investigating photodegradation and microbial degradation of tDOM, using low volumes, have not been suitable for lignin measurements before, but this can now open for more simpler experimental designs where lignin measurements can be included.

2.4 LINKING TDOM IN FRAM STRAIT TO ITS ORIGIN

Tracing lignin all the way from land into the oceans, can potentially help us to understand the impact of climate change on the carbon cycling, especially with increasing permafrost soil erosion into rivers and coastal waters (Biskaborn et al., 2019; Günther et al., 2015; Romanovsky et al., 2010; Smith et al., 2010).

As tDOM from Arctic rivers and coastal areas travel into the Arctic Ocean it will likely not be subject to a large amount of photodegradation during its travel towards the Fram Strait. This is due to relative high turbidity in shelf waters, caused by self-shading of photosensitive molecules, protection against direct sunlight in the open ocean because of the perennial sea-ice, and finally due to low irradiance in the Arctic region on annual scale (Amon et al. 2003). This can be supported by the findings by Benner et al. (2005), which showed that, based on the size distribution and composition of lignin phenols, transformation of tDOM appears to not be extensive in PSW throughout the Arctic Ocean. Kaiser et al. (2017) also showed that S/V

ratios and ratio between p-hydroxy (P) and V phenols (P/V) in fact can be used as tracers to study the mixing between rivers on the East Siberian shelf regions, thereby linking the tDOM to its river source. For Paper C, an attempt to link tDOM in the Fram Strait (78°50' N, 17° W to 7.5° E) to its source, based on measured (HPLC-DAD) and modelled (N-PLS on EEMs) lignin phenol ratios, was therefore carried out to see if it was possible to differentiate between water masses.

The measured and predicted S/V, C/V and P/V ratios for PSW in the Fram Strait, which is exported out of the Arctic Ocean with the East Greenland Current, was compared to the lignin phenol ratios found in Arctic rivers (Amon et al., 2012; Lobbes et al., 2000; Mann et al., 2016). The PSW with the highest concentration of TDLP11 showed signatures related to lignin-derived material associated with East Siberian rivers, particularly Lena and Yenisey. This is not suprising, since the Lena and Yenisey rivers are the two largest contributors of tDOM into the Arctic (Amon et al., 2012; Mann et al., 2016; Opsahl et al., 1999). Linking the DOM in PSW, exiting through the Fram Strait, to its source has been performed purely based on optical measurements as well (Amon et al., 2003; Gonçalves-Araujo et al., 2016; Granskog et al., 2012). Gonçalves-Araujo et al. (2016) demonstrated how fluorescence spectroscopy could be used to link the tDOM in PSW to two different sources, the Canada basin and the Eurasian basin. Combining lignin phenols and fluorescence spectroscopy could therefore aid each other in unravelling the origin of water masses in the Fram Strait.

For most of the seawater samples taken across the PSW of the Fram Strait, the lignin phenol ratios seemed to differ slightly from the ratios of major Arctic rivers. The deviation in lignin phenol ratios to their river source, in PSW, can be due to mixing between different sources of tDOM during transport on the shelfves as observed by Kaiser et al. (2017). So rather than using the Arctic rivers as reference, it may be more appropriate to compare lignin phenol ratios in the Fram Strait with endmembers obtained from locations farther out on the shelf, to determine their origin. However, the findings in Paper C, still adds to expand our understanding on the connectivity between the tDOM released by Arctic rivers and tDOM found in PSW in the Fram Strait.

The Atlantic waters (AW), imported to the Arctic Ocean through the Fram Strait with the West Spitsbergen Current, seemed to mainly carry one fingerprint, very different from the PSW. However, surprisingly another source of terrestrial material, different from the rest of the AW, was found in the most eastern side of the transect, close to Svalbard, from the predicted lignin phenol ratios. The uncovering a distinct water mass close to the coast of Svalbard, was not detected from bulk fluorescence measurements in the first place and shows how the use of N-PLS for prediction of lignin phenols indeed can reveal some new insights.

2.5 THE POTETIONAL FOR PREDICTING LIGNIN PHENOLS FROM BULK DOM FLUORESCENCE

Based on the lignin measurements in Paper C, an N-PLS prediction model was built to predict lignin phenol parameters from bulk fluorescence EEMs in the open ocean. The N-PLS showed to be able to predict TDLP11, S/V, C/V and P/V with r^2 values >0.54. The N-PLS model could however not predict Ad/Al (V) and Ad/Al (S). Predicting lignin phenol parameters in open ocean waters where TDLP11 concentrations are low (< 6 nM) has until now been challenging (Fichot et al., 2016). However, the results in Paper C, shows for that N-PLS modelling applied to EEMs maybe can help solve this challenge.

An interesting aspect of Fram Strait is the continuous monitoring, where fluorescence measurements have been collected over the last decade. Applying the N-PLS model to the EEMs from previous years can allow us to look into past TDLP11 concentrations in the Fram Strait. By applying the N-PLS to the EEMs from previous years, it was found that TDLP11 in the exported PSW has stayed more or less stable the last decade, especially if the variation for each year is taken into account (see Figure 17). However, from Figure 17 there still seems to be some kind of cycle, where TDPL11 increases and decreases between the years, and especially in year 2017 a sudden increase in lignin concentration seems to be predicted. The more or less stable lignin phenol export in the PSW fits with findings from Gonçalves-Araujo et al. (2020) where they found that the DOC exports also were reasonable constant over the last decade, especially in the first period.

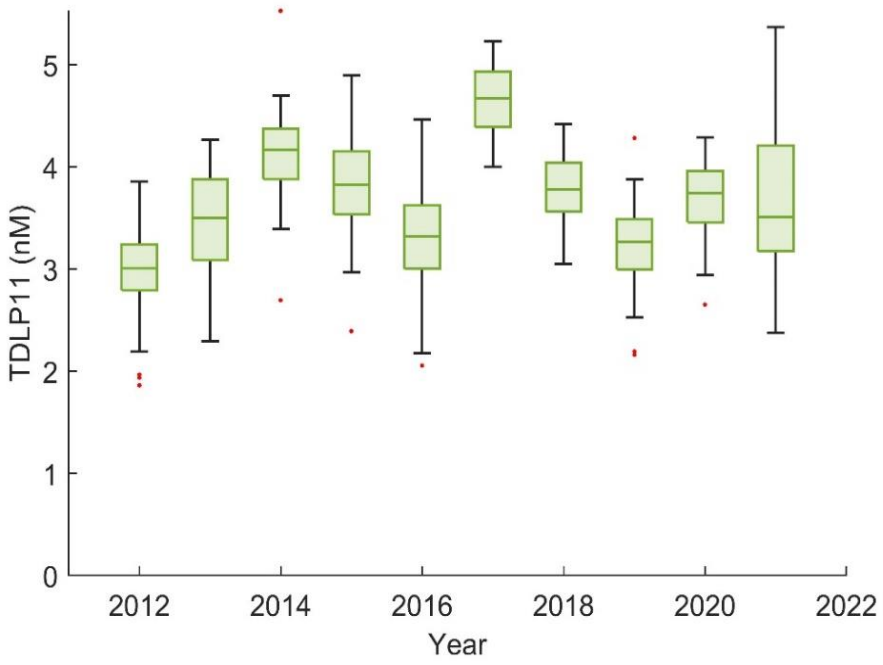


Figure 17: A whisker plot for TDLP across all 11 lignin phenols extrapolated by the developed N-PLS model from EEMs collected in Arctic surface waters (temperature<0 and salinity<32) in the Fram Strait over the last ten years (2012-2021). The red dots indicate outlier concentrations.

Fichot et al. 2016 showed how their model for prediction of TDLP9 could be used to predict to distrubtion of lignin phenols along a shelf transect in the Middle Altantic Bight. An idea with the developed N-PLS model in Paper C, would therefore be to try to predict the distribution of lignin-derived material across the shelf regions in the Arctic Ocean, using EEMs collected by other studies. This could help us to study the fate and diagenesis of lignin material from Arctic rivers in larger pan Arctic perspective. Finally, another interesting idea would also be to test the N-PLS model on EEMs from other ocean basins, where lignin phenol concentration has been measured as well, to see if the model can perform outside the Fram Strait.

The developed N-PLS model was further simplified, using only four excitation wavelengths with full emission scans. The large reduction in excitations wavelengths showed to perform nearly as good as the full size N-PLS model. Being reduced to only four excitation wavelengths means that this model can help to pave the way for lignin measurements from spectral in-situ sensors. These sensors are currently under development (Zielinski et al., 2018), but in the near future we could potentially have moorings, drones and Agrofloats submerged in the ocean with in-situ spectral sensors attached. Using the reduced N-PLS model on in-situ sensors could therefore allow us to measure lignin over a longer time period and for longer distances, not being limited to annually research expeditions to certain areas.

Our new proposed N-PLS approach for predicting lignin phenols concentrations based on EEMs can in a larger perspective have a significant positive impact on the environment and climate. By applying this model in the future, the need for chemicals (cupric oxidation and chromatography) and research expeditions on ships is reduced. The reduction in expeditions means less emission of greenhouse house gasses from ships, thereby reducing the overall carbon footprint of the research. Additionally, the reduced use of chemicals achieved by more efficient methods of data collection and analysis aligns with the principles of green chemistry, which focuses on minimizing the use of hazardous chemicals and reducing waste (Anastas & Eghbali, 2009). The reduction in emission and hazardeous chemical is also aligned with multiple aspects of the United Nation’s Sustainable Development Goals (United Nations, 2023). This innovative approach to studying dissolved lignin phenol concentrations therefore not only improves our understanding of the environment, but also has a direct positive effect on it by reducing emissions and promoting sustainable practices.

GENEREAL CONCLUSION

The research in my Ph.D. study attempts to understand the continuum of tDOM from its release from land areas encircling the Arctic Ocean and its fate in the ACZ to its exports through major Arctic gateways, such as Davis Strait and Fram Strait.

From my research, it was found that erosion of permafrost soils along Arctic coastlines will introduce tDOM with slightly different characters into the ACZ. The slight difference between soil types led to different biodegradability and thereby influenced the development of distinct marine microbial communities. Even though the tDOM from all three permafrost soil types was found be readily degraded and altered in quality by marine microbes, all permafrost soil types contained a larger fraction of refractory tDOM, which potentially will be exported to the open ocean.

My Ph.D. study found that the refractory fraction of tDOM escaping the ACZ, more specifically lignin-derived material, can be found in PSW being exported to the Atlantic Ocean through the Fram Strait. Based on lignin phenol ratios, it was found that the terrestrial material released from the Arctic rivers Lena and Yenisey may dominate the water masses with highest TDLP concentrations.

The thesis provides a new machine-learning approach for quantification of lignin phenols for HPLC-DAD that circumnavigate challenges with complex DOM samples and therefore improves sensitivity and specificity. The new method therefore provides scientists and laboratories with an option to measure lignin in the ocean using a simple HPLC-DAD setup, which seems to be competitive to MS. Addtionally, the new method reduces required seawater volumes needed for quantification, using HPLC-DAD, and therefore opens up for less sampling efforts on research expeditions.

Finally, does my Ph.D. study provide a tool, more specifically an N-PLS model, to predict lignin parameters from fluorescence measurements in the open ocean, more particular the Fram Strait. The developed N-PLS model will make it possible to extrapolate lignin phenol parameters onto larger spatial coverage and over longer timescales, since fluorescence measurements are easily obtained. Ultimately, the N-PLS model can be applied onto in-situ fluoresence sensors which can help us to monitor the effect of climate change on the ocean’s carbon budget more closely in in the future.

REFERENCE LIST

Aiken, G. (2014). Fluorescence and Dissolved Organic Matter. *Aquatic Organic Matter Fluorescence*, 35–74. https://doi.org/10.1017/CBO9781139045452.005

Akpan, E. I. (2019). Chemistry and Structure of Lignin. *Sustainable Lignin for Carbon Fibers: Principles, Techniques, and Applications*, 1–50. https://doi.org/10.1007/978-3-030-18792-7_1

Albinsson, B., Li, S., Lundquist, K., & Stomberg, R. (1999). The origin of lignin fluorescence. *Journal of Molecular Structure*, *508*(1–3), 19–27. https://doi.org/10.1016/S0022-2860(98)00913-2

Amigo, J. M., Skov, T., Bro, R., Coello, J., & Maspoch, S. (2008). Solving GC-MS problems with PARAFAC2. *TrAC Trends in Analytical Chemistry*, *27*(8), 714–725. https://doi.org/10.1016/J.TRAC.2008.05.011

Amon, R. M. W., & Budéus, G. (2003). Dissolved organic carbon distribution and origin in the Nordic Seas: Exchanges with the Arctic Ocean and the North Atlantic. *Journal of Geophysical Research*, *108*(C7), 3221. https://doi.org/10.1029/2002JC001594

Amon, R. M. W., Budéus, G., & Meon, B. (2003). Dissolved organic carbon distribution and origin in the Nordic Seas: Exchanges with the Arctic Ocean and the North Atlantic. *Journal of Geophysical Research: Oceans*, *108*(C7), 3221. https://doi.org/10.1029/2002JC001594

Amon, R. M. W., & Meon, B. (2004). The biogeochemistry of dissolved organic matter and nutrients in two large Arctic estuaries and potential implications for our understanding of the Arctic Ocean system. *Marine Chemistry*, *92*(1-4 SPEC. ISS.), 311–330. https://doi.org/10.1016/j.marchem.2004.06.034

Amon, R. M. W., Rinehart, A. J., Duan, S., Louchouarn, P., Prokushkin, A., Guggenberger, G., Bauch, D., Stedmon, C., Raymond, P. A., Holmes, R. M., McClelland, J. W., Peterson, B. J., Walker, S. A., & Zhulidov, A. v. (2012). Dissolved organic matter sources in large Arctic rivers. *Geochimica et Cosmochimica Acta*, *94*, 217–237. https://doi.org/10.1016/j.gca.2012.07.015

Anastas, P., & Eghbali, N. (2009). Green Chemistry: Principles and Practice. *Chemical Society Reviews*, *39*(1), 301–312. https://doi.org/10.1039/B918763B

Anderson, L. G., & Amon, R. M. W. (2015). DOM in the Arctic Ocean. *Biogeochemistry of Marine Dissolved Organic Matter: Second Edition*, 609–633. https://doi.org/10.1016/B978-0-12-405940-5.00014-5

Andrew, A. A., del Vecchio, R., Subramaniam, A., & Blough, N. v. (2013). Chromophoric dissolved organic matter (CDOM) in the Equatorial Atlantic Ocean: Optical properties and their relation to CDOM structure and source. *Marine Chemistry*, *148*, 33–43. https://doi.org/10.1016/J.MARCHEM.2012.11.001

Arellano, A. R., Bianchi, T. S., Hutchings, J. A., Shields, M. R., & Cui, X. (2018). Differential effects of solid-phase extraction resins on the measurement of dissolved lignin-phenols and organic matter composition in natural waters. *Limnology and Oceanography: Methods*, *16*(1), 22–34. https://doi.org/10.1002/lom3.10224

Arrigo, K. R., & van Dijken, G. L. (2011). Secular trends in Arctic Ocean net primary production. *Journal of Geophysical Research: Oceans*, *116*(9). https://doi.org/10.1029/2011JC007151

Baccolo, G., Quintanilla-Casas, B., Vichi, S., Augustijn, D., & Bro, R. (2021). From untargeted chemical profiling to peak tables – A fully automated AI driven approach to untargeted GC-MS. *TrAC Trends in Analytical Chemistry*, *145*, 116451. https://doi.org/10.1016/J.TRAC.2021.116451

Bai, X. M., Liu, T., Liu, D. L., & Wei, Y. J. (2018). Simultaneous determination of α-asarone and β-asarone in Acorus tatarinowii using excitation-emission matrix fluorescence coupled with chemometrics methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *191*, 195–202. https://doi.org/10.1016/J.SAA.2017.10.011

Balch, W. M., Drapeau, D. T., Bowler, B. C., & Huntington, T. G. (2012). Step-changes in the physical, chemical and biological characteristics of the Gulf of Maine, as documented by the GNATS time series. *Marine Ecology Progress Series*, *450*, 11–35. https://doi.org/10.3354/MEPS09555

Bao, H., Lee, T. Y., Huang, C., Feng, X., Dai, M., & Kao, S. J. (2015). Importance of Oceanian small mountainous rivers (SMRs) in global land-to-ocean output of lignin and modern biospheric carbon. *Scientific Reports*, *5*. https://doi.org/10.1038/srep16217

Beer. (1852). Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Annalen Der Physik*, *162*(5), 78–88. https://doi.org/10.1002/ANDP.18521620505

Benner, R. (2004). What happens to terrestrial organic matter in the ocean? *Marine Chemistry*, *92*(1-4 SPEC. ISS.), 307–310. https://doi.org/10.1016/j.marchem.2004.06.033

Benner, R., & Amon, R. M. W. (2015). The Size-Reactivity Continuum of Major Bioelements in the Ocean. *Annual Review of Marine Science*, *7*, 185–205. https://doi.org/10.1146/ANNUREV-MARINE-010213-135126

Benner, R., Louchouarn, P., & Amon, R. M. W. (2005). Terrigenous dissolved organic matter in the Arctic Ocean and its transport to surface and deep waters of the North Atlantic. *Global Biogeochemical Cycles*, *19*(2), 1–11. https://doi.org/10.1029/2004GB002398

Benner, R., & Opsahl, S. (2001). Molecular indicators of the sources and transformations of dissolved organic matter in the Mississippi river plume. *Organic Geochemistry*, *32*(4), 597–611. https://doi.org/10.1016/S0146-6380(00)00197-2

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., … Lantuit, H. (2019). Permafrost is warming at a global scale. *Nature Communications 2019 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41467-018-08240-4

Blanchet, M., Pringault, O., Panagiotopoulos, C., Lefèvre, D., Charrière, B., Ghiglione, J.-F., Fernandez, C., Aparicio, F. L., Marrasé, C., & Catala, P. (2017). When riverine dissolved organic matter (DOM) meets labile DOM in coastal waters: changes in bacterial community

activity and composition. *Aquatic Sciences-Research Across Boundaries*, *79*(1). https://doi.org/10.1007/s00027-016-0477-00

Bro, R. (1996). MULTIWAY CALIBRATION. MULTILINEAR PLS. *JOURNAL OF CHEMOMETRICS*, *10*, 47–61.

Bro, R., Andersson, C. A., & Kiers, H. A. L. (1999). PARAFAC2-PART II. MODELING CHROMATOGRAPHIC DATA WITH RETENTION TIME SHIFTS. *JOURNAL OF CHEMOMETRICS*, *13*, 295–309.

Bruhn, A. D., Stedmon, C. A., Comte, J., Matsuoka, A., Speetjens, N. J., Tanski, G., Vonk, J. E., & Sjöstedt, J. (2021). Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.640580

Bugg, T. D. H., Ahmad, M., Hardiman, E. M., & Rahmanpour, R. (2011). Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports*, *28*(12), 1883–1896. https://doi.org/10.1039/C1NP00042J

Caron, F., Elchuk, S., & Walker, Z. H. (1996). High-performance liquid chromatographic characterization of dissolved organic matter from low-level radioactive waste leachates. *Journal of Chromatography A*, *739*(1–2), 281–294. https://doi.org/10.1016/0021-9673(96)00077-5

Coble, P. G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry*, *51*(4), 325–346. https://doi.org/10.1016/0304-4203(95)00062-3

Coble, P. G. (2007). Marine Optical Biogeochemistry: The Chemistry of Ocean Color. *Chem. Rev.*, *107*, 402–418. https://doi.org/10.1021/cr050350

Cory, R. M., Crump, B. C., Dobkowski, J. A., & Kling, G. W. (2013). Surface exposure to sunlight stimulates CO2 release from permafrost soil carbon in the Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3429–3434. https://doi.org/10.1073/PNAS.1214104110

Cory, R. M., McKnight, D. M., Chin, Y. P., Miller, P., & Jaros, C. L. (2007). Chemical characteristics of fulvic acids from Arctic surface waters: Microbial contributions and photochemical transformations. *Journal of Geophysical Research: Biogeosciences*, *112*(4). https://doi.org/10.1029/2006JG000343

Cory, R. M., Ward, C. P., Crump, B. C., & Kling, G. W. (2014). Sunlight controls water column processing of carbon in arctic fresh waters. *Science*, *345*(6199), 925–928. https://doi.org/10.1126/SCIENCE.1253119

Cummins, P. M., Rochfort, K. D., & O'Connor, B. F. (2017). Ion-exchange chromatography: Basic principles and application. *Methods in Molecular Biology*, *1485*, 209–223. https://doi.org/10.1007/978-1-4939-6412-3_11

Dence, C. W., & Lin, S. Y. (1992). *Methods in Lignin Chemistry* (T. E. Timell, Ed.; pp. 3–19). Springer-Verlag . https://doi.org/10.1007/978-3-642-74065-7_1

Dey, S., Maiti, T. K., & Bhattacharyya, B. C. (1994). Production of some extracellular enzymes by a lignin peroxidase-producing brown rot fungus, Polyporus ostreiformis, and its comparative abilities for lignin degradation and dye decolorization. *Applied and Environmental Microbiology*, *60*(11), 4216. https://doi.org/10.1128/AEM.60.11.4216-4218.1994

Dinasquet, J., Kragh, T., Schrøter, M. L., Søndergaard, M., & Riemann, L. (2013). Functional and compositional succession of bacterioplankton in response to a gradient in bioavailable dissolved organic carbon. *Environmental Microbiology*, *15*(9), 2616–2628. https://doi.org/10.1111/1462-2920.12178

Dittmar, T., Koch, B., Hertkorn, N., & Kattner, G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods*, *6*(6), 230–235. https://doi.org/10.4319/LOM.2008.6.230

Fasham, M. J. R., Baliño, B. M., & Bowles, M. C. (2001). A new vision of ocean biochemistry after a decade of the joint global ocean flux study (JGOFS). *Ambio*, *30*(SPEC. ISS. 10), 4–30.

Fichot, C. G., Benner, R., Kaiser, K., Shen, Y., Amon, R. M. W., Ogawa, H., & Lu, C. J. (2016). Predicting dissolved lignin phenol concentrations in the coastal ocean from chromophoric dissolved organic matter (CDOM) absorption coefficients. *Frontiers in Marine Science*, *3*(FEB). https://doi.org/10.3389/fmars.2016.00007

Fichot, C. G., Kaiser, K., Hooker, S. B., Amon, R. M. W., Babin, M., Bélanger, S., Walker, S. A., & Benner, R. (2013). Pan-Arctic distributions of continental runoff in the Arctic Ocean. *Scientific Reports*, *3*. https://doi.org/10.1038/srep01053

Fritz, M., Vonk, J. E., & Lantuit, H. (2017). Collapsing Arctic coastlines. In *Nature Climate Change* (Vol. 7, Issue 1, pp. 6–7). Nature Publishing Group. https://doi.org/10.1038/nclimate3188

García, I., Ortiz, M. C., Sarabia, L., & Aldama, J. M. (2007). Validation of an analytical method to determine sulfamides in kidney by HPLC-DAD and PARAFAC2 with first-order derivative chromatograms. *Analytica Chimica Acta*, *587*(2), 222–234. https://doi.org/10.1016/J.ACA.2007.01.054

Gonçalves-Araujo, R., Granskog, M. A., Bracher, A., Azetsu-Scott, K., Dodd, P. A., & Stedmon, C. A. (2016). Using fluorescent dissolved organic matter to trace and distinguish the origin of Arctic surface waters. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP33978

Gonçalves-Araujo, R., Stedmon, C. A., de Steur, L., Osburn, C. L., & Granskog, M. A. (2020). A Decade of Annual Arctic DOC Export With Polar Surface Water in the East Greenland Current. *Geophysical Research Letters*, *47*(20). https://doi.org/10.1029/2020GL089686

Goñi, M. A., & Hedges, J. I. (1995). Sources and reactivities of marine-derived organic matter in coastal sediments as determined by alkaline CuO oxidation. *Geochimica et Cosmochimica Acta*, *59*(14), 2965–2981. https://doi.org/10.1016/0016-7037(95)00188-3

Granskog, M. A., Stedmon, C. A., Dodd, P. A., Amon, R. M. W., Pavlov, A. K., de Steur, L., & Hansen, E. (2012). Characteristics of colored dissolved organic matter (CDOM) in the Arctic outflow in the Fram Strait: Assessing the changes and fate of terrigenous CDOM in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *117*(12). https://doi.org/10.1029/2012JC008075

Green, S. A., & Blough, N. v. (1994). Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters. *Limnology and Oceanography*, *39*(8), 1903–1916. https://doi.org/10.4319/LO.1994.39.8.1903

Grunert, B. K., Tzortziou, M., Neale, P., Menendez, A., & Hernes, P. (2021). DOM degradation by light and microbes along the Yukon River-coastal ocean continuum. *Scientific Reports*, *11*(1), 10236. https://doi.org/10.1038/S41598-021-89327-9

Günther, F., Overduin, P. P., Yakshina, I. A., Opel, T., Baranskaya, A. v., & Grigoriev, M. N. (2015). Observing Muostakh disappear: Permafrost thaw subsidence and erosion of a ground-ice-rich Island in response to arctic summer warming and sea ice reduction. *Cryosphere*, *9*(1), 151–178. https://doi.org/10.5194/TC-9-151-2015

Hansell, D. A. (2013). Recalcitrant dissolved organic carbon fractions. *Annual Review of Marine Science*, *5*, 421–445. https://doi.org/10.1146/ANNUREV-MARINE-120710-100757

Hansell, D. A., Carlson, C. A., Repeta, D. J., & Schlitzer, R. (2009). Dissolved organic matter in the ocean a controversy stim ulates new insights. *Oceanography*, *22*(SPL.ISS. 4), 202–211. https://doi.org/10.5670/OCEANOG.2009.109

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Hawkes, J. A., Dittmar, T., Patriarca, C., Tranvik, L., & Bergquist, J. (2016). Evaluation of the Orbitrap Mass Spectrometer for the Molecular Fingerprinting Analysis of Natural Dissolved Organic Matter. *Analytical Chemistry*, *88*(15), 7698–7704. https://doi.org/10.1021/ACS.ANALCHEM.6B01624/SUPPL_FILE/AC6B01624_SI_003.TXT

Hawkes, J. A., Sjöberg, P. J. R., Sjöberg, S., Bergquist, J., & Tranvik, L. J. (2019). *Complexity of dissolved organic matter in the molecular size dimension: insights from coupled size exclusion chromatography electrospray ionisation mass spectrometry †*. https://doi.org/10.1039/c8fd00222c

Hedges, J. I. (1992). Global biogeochemical cycles: progress and problems. *Marine Chemistry*, *39*(1–3), 67–93. https://doi.org/10.1016/0304-4203(92)90096-S

Hedges, J. I. (2002). Why Dissolved Organics Matter. *Biogeochemistry of Marine Dissolved Organic Matter*, 1–33. https://doi.org/10.1016/B978-012323841-2/50003-8

Hedges, J. I., Blanchette, R. A., Weliky, K., & Devol, A. H. (1988). Effects of fungal degradation on the CuO oxidation products of lignin: A controlled laboratory study. *Geochimica et Cosmochimica Acta*, *52*(11), 2717–2726. https://doi.org/10.1016/0016-7037(88)90040-3

Hedges, J. I., & Ertel, J. R. (1982). Characterization of Lignin by Gas Capillary Chromatography of Cupric Oxide Oxidation Products. *Analytical Chemistry*, *54*(2), 174–178. https://doi.org/10.1021/AC00239A007

Hedges, J. I., Keil, R. G., & Benner, R. (1997). What Happens to Terrestrial Organic Matter in the Ocean? *Organic Geochemistry*, *27*(5–6), 195. https://doi.org/https://doi.org/10.1016/S0146-6380(97)00066-1

Hedges, J. I., & Mann, D. C. (1979). The characterization of plant tissues by their lignin oxidation products. *Geochimica et Cosmochimica Acta*, *43*(11), 1803–1807. https://doi.org/10.1016/0016-7037(79)90028-0

Helms, J. R., Stubbins, A., Ritchie, J. D., Minor, E. C., Kieber, D. J., & Mopper, K. (2008). Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter. *Limnology and Oceanography*, *53*(3), 955–969. https://doi.org/10.4319/LO.2008.53.3.0955

Herlemann, D. P. R., Manecki, M., Meeske, C., Pollehne, F., Labrenz, M., Schulz-Bull, D., Dittmar, T., & Jürgens, K. (2014). Uncoupling of bacterial and terrigenous dissolved organic matter dynamics in decomposition experiments. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0093945

Hernes, P. J., & Benner, R. (2002). Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean. In *Deep-Sea Research I* (Vol. 49).

Hernes, P. J., & Benner, R. (2003). Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. *Journal of Geophysical Research: Oceans*, *108*(9). https://doi.org/10.1029/2002jc001421

Hernes, P. J., Bergamaschi, B. A., Eckard, R. S., Hernes, P. J., Bergamaschi, B. A., & Eckard, R. S. (2006). Potential utility of optical proxies for lignin in DOM. *AGUFM*, *2006*, B11C-1041. https://ui.adsabs.harvard.edu/abs/2006AGUFM.B11C1041H/abstract

Hernes, P. J., Bergamaschi, B. A., Eckard, R. S., & Spencer, R. G. M. (2009). Fluorescence-based proxies for lignin in freshwater dissolved organic matter. *Journal of Geophysical Research: Biogeosciences*, *114*(4). https://doi.org/10.1029/2009JG000938

Hernes, P. J., Spencer, R. G. M., Dyda, R. Y., Pellerin, B. A., Bachand, P. A. M., & Bergamaschi, B. A. (2008). The role of hydrologic regimes on dissolved organic carbon composition in an agricultural watershed. *Geochimica et Cosmochimica Acta*, *72*(21), 5266–5277. https://doi.org/10.1016/J.GCA.2008.07.031

Holmes, R. M., McClelland, J. W., Peterson, B. J., Tank, S. E., Bulygina, E., Eglinton, T. I., Gordeev, V. v., Gurtovaya, T. Y., Raymond, P. A., Repeta, D. J., Staples, R., Striegl, R. G., Zhulidov, A. v., & Zimov, S. A. (2012). Seasonal and Annual Fluxes of Nutrients and Organic Matter from Large Rivers to the Arctic Ocean and Surrounding Seas. *Estuaries and Coasts*, *35*(2), 369–382. https://doi.org/10.1007/s12237-011-9386-6

Holmes, R. M., Mcclelland, J. W., Raymond, P. A., Frazer, B. B., Peterson, B. J., Stieglitz, M., Mcclelland, W., Raymond, P. A., Frazer, B. B., Peterson, B. J., & Stieglitz, M. (2008). Lability of DOC transported by Alaskan rivers to the Arctic Ocean. *Geophysical Research Letters*, *35*(3). https://doi.org/10.1029/2007GL032837

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, *2*(3), 211–228. https://doi.org/10.1002/CEM.1180020306

Jones, E. P., Rudels, B., & Anderson, L. G. (1995). Deep waters of the Arctic Ocean: origins and circulation. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*(5), 737–760. https://doi.org/10.1016/0967-0637(95)00013-V

Kaiser, K., & Benner, R. (2012). Characterization of lignin by gas chromatography and mass spectrometry using a simplified CuO oxidation method. *Analytical Chemistry*, *84*(1), 459–464. https://doi.org/10.1021/ac202004r

Kaiser, K., Benner, R., & Amon, R. M. W. (2017). The fate of terrigenous dissolved organic carbon on the Eurasian shelves and export to the North Atlantic. *Journal of Geophysical Research: Oceans*, *122*(1), 4–22. https://doi.org/10.1002/2016JC012380

Kiers, H. A. L., ten Berge, J. M. F., & Bro, R. (1999). PARAFAC2-PART I. A DIRECT FITTING ALGORITHM FOR THE PARAFAC2 MODEL. *J. Chemometrics*, *13*, 275–294. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4

Koch, B. P., & Dittmar, T. (2006). From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter. *EPIC3Rapid Communications in Mass Spectrometry, 20(5), Pp. 926-932*, *20*(5), 926–932. https://doi.org/10.1002/RCM.2386

Koch, B. P., Ludwichowski, K. U., Kattner, G., Dittmar, T., & Witt, M. (2008). Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Marine Chemistry*, *3–4*(111), 233–241. https://doi.org/10.1016/J.MARCHEM.2008.05.008

Kumar, K., & Mishra, A. K. (2012). Quantification of ethanol in ethanol-petrol and biodiesel in biodiesel-diesel blends using fluorescence spectroscopy and multivariate methods. *Journal of Fluorescence*, *22*(1), 339–347. https://doi.org/10.1007/S10895-011-0964-6/TABLES/5

Lewis, N. G., & Yamamoto, E. (1990). LIGNIN: OCCURRENCE, BIOGENESIS AND BIODEGRADATION. *Annu. Rev. Plant Physiol. Plant Mol. Biol*, *41*, 455–496.

Lin, Y., Li, X. K., Yan, J. Y., Lei, Y. L., Chen, J., & Li, B. Q. (2022). EEFM combined with N-PLS method for the quantitative determination of psoralen in cosmetics. *Chemical Papers*. https://doi.org/10.1007/S11696-022-02262-5

Lobbes, J. M., Fitznar, H. P., & Kattner, G. (1999). High-performance liquid chromatography of lignin-derived phenols in environmental samples with diode array detection. *Analytical Chemistry*, *71*(15), 3008–3012. https://doi.org/10.1021/ac981175i

Lobbes, J. M., Fitznar, H. P., & Kattner, G. (2000). Biogeochemical characteristics of dissolved and particulate organic matter in Russian rivers entering the Arctic Ocean. *Geochimica et Cosmochimica Acta*, *64*(17), 2973–2983. https://doi.org/10.1016/S0016-7037(00)00409-9

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Louchouam, P., Opsahl, S., & Benner, R. (2000). Isolation and Quantification of Dissolved Lignin from Natural Waters Using Solid-Phase Extraction and GC/MS. *Analytical Chemistry*, *72*(13), 2780–2787. https://doi.org/10.1021/AC9912552

Lu, J., Wang, M., Zhang, X., Heyden, A., & Wang, F. (2016). β-O-4 Bond Cleavage Mechanism for Lignin Model Compounds over Pd Catalysts Identified by Combination of First-Principles Calculations and Experiments. *ACS Catalysis*, *6*(8), 5589–5598. https://doi.org/10.1021/ACSCATAL.6B00502

Lu, Y., Lu, Y. C., Hu, H. Q., Xie, F. J., Wei, X. Y., & Fan, X. (2017). Structural characterization of lignin and its degradation products with spectroscopic methods. *Journal of Spectroscopy*, *2017*. https://doi.org/10.1155/2017/8951658

Manizza, M., Follows, M. J., Dutkiewicz, S., McClelland, J. W., Menemenlis, D., Hill, C. N., Townsend-Small, A., & Peterson, B. J. (2009). Modeling transport and fate of riverine dissolved organic carbon in the Arctic Ocean. *Global Biogeochemical Cycles*, *23*(4). https://doi.org/10.1029/2008GB003396

Mann, P. J., Eglinton, T. I., McIntyre, C. P., Zimov, N., Davydova, A., Vonk, J. E., Holmes, R. M., & Spencer, R. G. M. (2015). Utilization of ancient permafrost carbon in headwaters of Arctic fluvial networks. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms8856

Mann, P. J., Spencer, R. G. M., Hernes, P. J., Six, J., Aiken, G. R., Tank, S. E., McClelland, J. W., Butler, K. D., Dyda, R. Y., & Holmes, R. M. (2016). Pan-arctic trends in terrestrial dissolved organic matter from optical measurements. *Frontiers in Earth Science*, *4*. https://doi.org/10.3389/feart.2016.00025

Marschner, B., & Kalbitz, K. (2003). Controls of bioavailability and biodegradability of dissolved organic matter in soils. *Geoderma*, *113*(3–4), 211–235. https://doi.org/10.1016/S0016-7061(02)00362-2

Matero, S., Poutiainen, S., Leskinen, J., Järvinen, K., Ketolainen, J., Poso, A., & Reinikainen, S. P. (2010). Estimation of granule size distribution for batch fluidized bed granulation process using acoustic emission and N-way PLS. *Journal of Chemometrics*, *24*(7–8), 464–471. https://doi.org/10.1002/CEM.1269

Mathis, J. T., Hansell, D. A., Kadko, D., Bates, N. R., & Cooper, L. W. (2007). Determining net dissolved organic carbon production in the hydrographically complex western Arctic Ocean. *Limnology and Oceanography*, *52*(5), 1789–1799. https://doi.org/10.4319/LO.2007.52.5.1789

McClelland, J. W., Holmes, R. M., Dunton, K. H., & Macdonald, R. W. (2012). The Arctic Ocean Estuary. *Estuaries and Coasts*, *35*(2), 353–368. https://doi.org/10.1007/S12237-010-9357-3/FIGURES/8

Mills, G. L., McFadden, E., & Quinn, J. G. (1987). Chromatographic studies of dissolved organic matter and copper-organic complexes isolated from estuarine waters. *Marine Chemistry*, *20*(4), 313–325. https://doi.org/10.1016/0304-4203(87)90065-X

Monod, J. (1950). LA TECHNIQUE DE CULTURE CONTINUE THÉORIE ET APPLICATIONS. *Annales d'Institut Pasteur*, *79*, 390–410. https://doi.org/10.1016/B978-0-12-460482-7.50023-3

Monties, B., & Fukushima, K. (2001). Occurrence, Function and Biosynthesis of Lignins. *Biopolymers Online*. https://doi.org/10.1002/3527600035.BPOL1001

Mopper, K., Stubbins, A., Ritchie, J. D., Bialk, H. M., & Hatcher, P. G. (2007). Advanced instrumental approaches for characterization of marine dissolved organic matter: Extraction techniques, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Chemical Reviews*, *107*(2), 419–442. https://doi.org/10.1021/CR050359B

Moriarty, D. J. W. (1986). *Measurement of Bacterial Growth Rates in Aquatic Systems from Rates of Nucleic Acid Synthesis*. 245–292. https://doi.org/10.1007/978-1-4757-0611-6_6

Müller, O., Seuthe, L., Bratbak, G., & Paulsen, M. L. (2018). Bacterial response to permafrost derived organic matter input in an Arctic Fjord. *Frontiers in Marine Science*, *5*(AUG). https://doi.org/10.3389/fmars.2018.00263

Murphy, K. R., Hambly, A., Singh, S., Henderson, R. K., Baker, A., Stuetz, R., & Khan, S. J. (2011). Organic matter fluorescence in municipal water recycling schemes: toward a unified PARAFAC model. *Environmental Science & Technology*, *45*(7), 2909–2916. https://doi.org/10.1021/ES103015E

Murphy, K. R., Stedmon, C. A., Graeber, D., & Bro, R. (2013). Fluorescence spectroscopy and multi-way techniques. PARAFAC. *Analytical Methods*, *5*(23), 6557–6566. https://doi.org/10.1039/C3AY41160E

Nielsen, D. M., Pieper, P., Barkhordarian, A., Overduin, P., Ilyina, T., Brovkin, V., Baehr, J., & Dobrynin, M. (2022). Increase in Arctic coastal erosion and its sensitivity to warming in the twenty-first century. *Nature Climate Change 2022 12:3*, *12*(3), 263–270. https://doi.org/10.1038/s41558-022-01281-0

Norgren, M., & Edlund, H. (2014). Lignin: Recent advances and emerging applications. *Current Opinion in Colloid & Interface Science*, *19*(5), 409–416. https://doi.org/10.1016/J.COCIS.2014.08.004

Novick, A., & Szilard, L. (1950). Experiments with the Chemostat on spontaneous mutations of bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *36*(12), 708–719. https://doi.org/10.1073/PNAS.36.12.708

Opsahl, S., & Benner, R. (1995). Early diagenesis of vascular plant tissues: Lignin and cutin decomposition and biogeochemical implications. *Geochimica et Cosmochimica Acta*, *59*(23), 4889–4904. https://doi.org/10.1016/0016-7037(95)00348-7

Opsahl, S., & Benner, R. (1997). Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature 1997 386:6624*, *386*(6624), 480–482. https://doi.org/10.1038/386480a0

Opsahl, S., & Benner, R. (1998). Photochemical reactivity of dissolved lignin in river and ocean waters. *Limnology and Oceanography*, *43*(6), 1297–1304. https://doi.org/10.4319/LO.1998.43.6.1297

Opsahl, S., Benner, R., & Amon, R. M. W. (1999). Major flux of terrigenous dissolved organic matter through the Arctic Ocean. *Limnology and Oceanography*, *44*(8), 2017–2023. https://doi.org/10.4319/LO.1999.44.8.2017

Osburn, C. L., Boyd, T. J., Montgomery, M. T., Bianchi, T. S., Coffin, R. B., & Paerl, H. W. (2016). Optical proxies for terrestrial dissolved organic matter in estuaries and coastal waters. *Frontiers in Marine Science*, *2*(JAN), 127. https://doi.org/10.3389/FMARS.2015.00127/BIBTEX

Osburn, C. L., Retamal, L., & Vincent, W. F. (2009). Photoreactivity of chromophoric dissolved organic matter transported by the Mackenzie River to the Beaufort Sea. *Marine Chemistry*, *115*(1–2), 10–20. https://doi.org/10.1016/J.MARCHEM.2009.05.003

Parlanti, E., Morin, B., & Vacher, L. (2002). Combined 3D-spectrofluorometry, high performance liquid chromatography and capillary electrophoresis for the characterization of dissolved organic matter in natural waters. *Organic Geochemistry*, *33*(3), 221–236. https://doi.org/10.1016/S0146-6380(01)00154-1

Patriarca, C., Bergquist, J., Sjöberg, P. J. R., Tranvik, L., & Hawkes, J. A. (2018). Online HPLC-ESI-HRMS Method for the Analysis and Comparison of Different Dissolved Organic Matter Samples. *Environmental Science and Technology*, *52*(4), 2091–2099. https://doi.org/10.1021/ACS.EST.7B04508/ASSET/IMAGES/LARGE/ES-2017-04508X_0004.JPEG

Paulsen, M. L., Nielsen, S. E. B., Müller, O., Møller, E. F., Stedmon, C. A., Juul-Pedersen, T., Markager, S., Sejr, M. K., Huertas, A. D., Larsen, A., & Middelboe, M. (2017). Carbon bioavailability in a high arctic fjord influenced by glacial Meltwater, NE Greenland. *Frontiers in Marine Science*, *4*(JUN). https://doi.org/10.3389/fmars.2017.00176

Perminova, I. v., Dubinenkov, I. v., Kononikhin, A. S., Konstantinov, A. I., Zherebker, A. Y., Andzhushev, M. A., Lebedev, V. A., Bulygina, E., Holmes, R. M., Kostyukevich, Y. I., Popov, I. A., & Nikolaev, E. N. (2014). Molecular mapping of sorbent selectivities with respect to isolation of arctic dissolved organic matter as measured by fourier transform mass spectrometry. *Environmental Science and Technology*, *48*(13), 7461–7468. https://doi.org/10.1021/ES5015423/SUPPL_FILE/ES5015423_SI_001.PDF

Petersen, M. A., & Bro, R. (2018). PARADISe - a ground-breaking tool to treat complex GC-MS datasets. In B. Siegmund & E. Leitner (Eds.), *Flavour Science: Proceedings of the XV*

Weurman Flavour Research Symposium (Vol. 350, pp. 421–426). Technischen Universität Graz.

Petras, D., Koester, I., Silva, R. da, Stephens, B. M., Haas, A. F., Nelson, C. E., Kelly, L. W., Aluwihare, L. I., & Dorrestein, P. C. (2017). High-resolution liquid chromatography tandem mass spectrometry enables large scale molecular characterization of dissolved organic matter. *Frontiers in Marine Science*, *4*(DEC), 405. https://doi.org/10.3389/FMARS.2017.00405/BIBTEX

Ramachandra, M., Crawford, D. L., Hertel, G., Ramachandra, (M, Crawford, D. L., & Pometto, A. L. (1988). Characterization of an extracellular lignin peroxidase of the lignocellulolytic actinomycete Streptomyces viridosporus. *Applied and Environmental Microbiology*, *54*(12), 3057–3063. https://doi.org/10.1128/AEM.54.12.3057-3063.1988

Rantanen, M., Karpechko, A. Yu., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment 2022 3:1*, *3*(1), 1–10. https://doi.org/10.1038/s43247-022-00498-3

Raymond, P. A., McClelland, J. W., Holmes, R. M., Zhulidov, A. v., Mull, K., Peterson, B. J., Striegl, R. G., Aiken, G. R., & Gurtovaya, T. Y. (2007). Flux and age of dissolved organic carbon exported to the Arctic Ocean: A carbon isotopic study of the five largest arctic rivers. *Global Biogeochemical Cycles*, *21*(4). https://doi.org/10.1029/2007GB002934

Raymond, P. A., & Spencer, R. G. M. (2015). Riverine DOM. In *Biogeochemistry of Marine Dissolved Organic Matter: Second Edition* (pp. 509–533). Elsevier Inc. https://doi.org/10.1016/B978-0-12-405940-5.00011-X

Reuter, H., Gensel, J., Elvert, M., & Zak, D. (2017). Direct Analysis of Lignin Phenols in Freshwater Dissolved Organic Matter. *Analytical Chemistry*, *89*(24), 13449–13457. https://doi.org/10.1021/acs.analchem.7b03729

Romankevich, E. A., Vetrov, A. A., & Peresypkin, V. I. (2009). Organic matter of the World Ocean. *Russian Geology and Geophysics*, *50*(4), 299–307. https://doi.org/10.1016/J.RGG.2009.03.013

Romanovsky, V. E., Smith, S. L., & Christiansen, H. H. (2010). Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007–2009: a synthesis. *Permafrost and Periglacial Processes*, *21*(2), 106–116. https://doi.org/10.1002/PPP.689

Rudels, B. (2015). Arctic Ocean circulation, processes and water masses: A description of observations and ideas with focus on the period prior to the International Polar Year 2007-2009. *Progress in Oceanography*, *132*, 22–67. https://doi.org/10.1016/J.POCEAN.2013.11.006

Rudels, B., Anderson, L. G., & Jones, E. P. (1996). Formation and evolution of the surface mixed layer and halocline of the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *101*(C4), 8807–8821. https://doi.org/10.1029/96JC00143

Rudels, B., Jones, E. P., Erson, L. G., & Kattner, G. (1994). On the intermediate depth waters of the Arctic ocean. *EPIC3AGU Geophysical Monograph, 85, Pp. 33-46*.

Sandron, S., Davies, N. W., Wilson, R., Cardona, A. R., Haddad, P. R., Nesterenko, P. N., & Paull, B. (2018). Fractionation of Dissolved Organic Matter on Coupled Reversed-Phase Monolithic Columns and Characterisation Using Reversed-Phase Liquid Chromatography-High Resolution Mass Spectrometry. *Chromatographia*, *81*(2), 203–213. https://doi.org/10.1007/S10337-017-3324-0/FIGURES/9

Schauer, U., Muench, R. D., Rudels, B., & Timokhov, L. (1997). Impact of eastern Arctic shelf waters on the Nansen Basin intermediate layers. *Journal of Geophysical Research: Oceans*, *102*(C2), 3371–3382. https://doi.org/10.1029/96JC03366

Schlünz, B., & Schneider, R. R. (2000). Transport of terrestrial organic carbon to the oceans by rivers: re-estimating flux- and burial rates. *International Journal of Earth Sciences 2000 88:4*, *88*(4), 599–606. https://doi.org/10.1007/S005310050290

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., & Vonk, J. E. (2015). Climate change and the permafrost carbon feedback. *Nature 2015 520:7546*, *520*(7546), 171–179. https://doi.org/10.1038/NATURE14338

Sholkovitz, E. R. (1976). Flocculation of dissolved organic and inorganic matter during the mixing of river water and seawater. *Geochimica et Cosmochimica Acta*, *40*(7), 831–845. https://doi.org/10.1016/0016-7037(76)90035-1

Sipler, R. E., Kellogg, C. T. E., Connelly, T. L., Roberts, Q. N., Yager, P. L., & Bronk, D. A. (2017). Microbial community response to terrestrially derived dissolved organic matter in the coastal Arctic. *Frontiers in Microbiology*, *8*(JUN). https://doi.org/10.3389/fmicb.2017.01018

Skov, T., & Bro, R. (2008). Solving fundamental problems in chromatographic analysis. *Analytical and Bioanalytical Chemistry*, *390*(1), 281–285. https://doi.org/10.1007/S00216-007-1618-Z

Smith, S. L., Romanovsky, V. E., Lewkowicz, A. G., Burn, C. R., Allard, M., Clow, G. D., Yoshikawa, K., & Throop, J. (2010). Thermal state of permafrost in North America: a contribution to the international polar year. *Permafrost and Periglacial Processes*, *21*(2), 117–135. https://doi.org/10.1002/PPP.690

Spencer, R. G. M., Aiken, G. R., Butler, K. D., Dornblaser, M. M., Striegl, R. G., & Hernes, P. J. (2009). Utilizing chromophoric dissolved organic matter measurements to derive export and reactivity of dissolved organic carbon exported to the Arctic Ocean: A case study of the Yukon River, Alaska. *Geophysical Research Letters*, *36*(6). https://doi.org/10.1029/2008GL036831

Spencer, R. G. M., Aiken, G. R., Dyda, R. Y., Butler, K. D., Bergamaschi, B. A., & Hernes, P. J. (2010). Comparison of XAD with other dissolved lignin isolation techniques and a compilation of analytical improvements for the analysis of lignin in aquatic settings. *Organic Geochemistry*, *41*(5), 445–453. https://doi.org/10.1016/j.orggeochem.2010.02.004

Spencer, R. G. M., Aiken, G. R., Wickland, K. P., Striegl, R. G., & Hernes, P. J. (2008). Seasonal and spatial variability in dissolved organic matter quantity and composition from the Yukon River basin, Alaska. *Global Biogeochemical Cycles*, *22*(4). https://doi.org/10.1029/2008GB003231

Spencer, R. G. M., Mann, P. J., Dittmar, T., Eglinton, T. I., McIntyre, C., Holmes, R. M., Zimov, N., & Stubbins, A. (2015). Detecting the signature of permafrost thaw in Arctic rivers. *Geophysical Research Letters*, *42*(8), 2830–2835. https://doi.org/10.1002/2015GL063498

Spencer, R. G. M., Stubbins, A., Hernes, P. J., Baker, A., Mopper, K., Aufdenkampe, A. K., Dyda, R. Y., Mwamba, V. L., Mangangu, A. M., Wabakanghanzi, J. N., & Six, J. (2009). Photochemical degradation of dissolved organic matter and dissolved lignin phenols from the Congo River. *Journal of Geophysical Research: Biogeosciences*, *114*(G3), 3010. https://doi.org/10.1029/2009JG000968

Stedmon, C. A., Amon, R. M. W., Rinehart, A. J., & Walker, S. A. (2011). The supply and characteristics of colored dissolved organic matter (CDOM) in the Arctic Ocean: Pan Arctic trends and differences. *Marine Chemistry*, *124*(1–4), 108–118. https://doi.org/10.1016/j.marchem.2010.12.007

Stedmon, C. A., & Markager, S. (2001). The optics of chromophoric dissolved organic matter (CDOM) in the Greenland Sea: An algorithm for differentiation between marine and terrestrially derived organic matter. *Limnology and Oceanography*, *46*(8), 2087–2093. https://doi.org/10.4319/LO.2001.46.8.2087

Stedmon, C. A., Markager, S., & Bro, R. (2003). Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. *Marine Chemistry*, *82*(3–4), 239–254. https://doi.org/10.1016/S0304-4203(03)00072-0

Stedmon, C. A., & Nelson, N. B. (2015). The Optical Properties of DOM in the Ocean. *Biogeochemistry of Marine Dissolved Organic Matter: Second Edition*, 481–508. https://doi.org/10.1016/B978-0-12-405940-5.00010-8

Stewart, F. J., Dalsgaard, T., Young, C. R., Thamdrup, B., Revsbech, N. P., Ulloa, O., Canfield, D. E., & DeLong, E. F. (2012). Experimental Incubations Elicit Profound Changes in Community Transcription in OMZ Bacterioplankton. *PLOS ONE*, *7*(5), e37118. https://doi.org/10.1371/JOURNAL.PONE.0037118

Tanski, G., Couture, N., Lantuit, H., Eulenburg, A., & Fritz, M. (2016). Eroding permafrost coasts release low amounts of dissolved organic carbon (DOC) from ground ice into the nearshore zone of the Arctic Ocean. *Global Biogeochemical Cycles*, *30*(7), 1054–1068. https://doi.org/10.1002/2015GB005337

Tanski, G., Lantuit, H., Ruttor, S., Knoblauch, C., Radosavljevic, B., Strauss, J., Wolter, J., Irrgang, A. M., Ramage, J., & Fritz, M. (2017). Transformation of terrestrial organic matter along thermokarst-affected permafrost coasts in the Arctic. *Science of the Total Environment*, *581–582*, 434–447. https://doi.org/10.1016/j.scitotenv.2016.12.152

Tanski, G., Wagner, D., Knoblauch, C., Fritz, M., Sachs, T., & Lantuit, H. (2019). Rapid CO2 Release From Eroding Permafrost in Seawater. *Geophysical Research Letters*, *46*(20), 11244–11252. https://doi.org/10.1029/2019GL084303

Terhaar, J., Lauerwald, R., Regnier, P., Gruber, N., & Bopp, L. (2021). Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion. *Nature Communications 2021 12:1*, *12*(1), 1–10. https://doi.org/10.1038/s41467-020-20470-z

Thingstad, T. F., Bellerby, R. G. J., Bratbak, G., Børsheim, K. Y., Egge, J. K., Heldal, M., Larsen, A., Neill, C., Nejstgaard, J., Norland, S., Sandaa, R. A., Skjoldal, E. F., Tanaka, T., Thyrhaug, R., & Töpper, B. (2008). Counterintuitive carbon-to-nutrient coupling in an Arctic pelagic ecosystem. *Nature 2008 455:7211*, *455*(7211), 387–390. https://doi.org/10.1038/NATURE07235

Tolbert, A., Akinosho, H., Khunsupat, R., Naskar, A. K., & Ragauskas, A. J. (2014). Characterization and analysis of the molecular weight of lignin for biorefining studies. *Biofuels, Bioproducts and Biorefining*, *8*(6), 836–856. https://doi.org/10.1002/BBB.1500

Traving, S. J., Rowe, O., Jakobsen, N. M., Sørensen, H., Dinasquet, J., Stedmon, C. A., Andersson, A., & Riemann, L. (2017). The effect of increased loads of dissolved organic matter on estuarine microbial community composition and function. *Frontiers in Microbiology*, *8*(MAR). https://doi.org/10.3389/fmicb.2017.00351

Tucker, L. R. (1951). A method for synthesis of factor analysis studie. *Psychometrika*, *16*(3), 121–136. https://doi.org/10.21236/AD0047524

United Nations. (2023). *United Nations: Sustainable Development*. Sustainable Development Goals (SDGs). https://sdgs.un.org/goals <acessed March 5, 2023>

Vivó-Truyols, G., Torres-Lapasió, J. R., van Nederkassel, A. M., vander Heyden, Y., & Massart, D. L. (2005). Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: Part I: Peak detection. *Journal of Chromatography A*, *1096*(1–2), 133–145. https://doi.org/10.1016/J.CHROMA.2005.03.092

Vonk, J. E., Mann, P. J., Davydov, S., Davydova, A., Spencer, R. G. M., Schade, J., Sobczak, W. v., Zimov, N., Zimov, S., Bulygina, E., Eglinton, T. I., & Holmes, R. M. (2013). High biolability of ancient permafrost carbon upon thaw. *Geophysical Research Letters*, *40*(11), 2689–2693. https://doi.org/10.1002/grl.50348

Vosough, M., & Salemi, A. (2011). Exploiting second-order advantage using PARAFAC2 for fast HPLC-DAD quantification of mixture of aflatoxins in pistachio nuts. *Food Chemistry*, *127*(2), 827–833. https://doi.org/10.1016/J.FOODCHEM.2011.01.017

Walker, S. A., Amon, R. M. W., Stedmon, C., Duan, S., & Louchouarn, P. (2009). The use of PARAFAC modeling to trace terrestrial dissolved organic matter and fingerprint water masses in coastal Canadian Arctic surface waters. *Journal of Geophysical Research: Biogeosciences*, *114*(4). https://doi.org/10.1029/2009JG000990

Wegner, C., Bennett, K. E., de Vernal, A., Forwick, M., Fritz, M., Heikkilä, M., Łącka, M., Lantuit, H., Laska, M., Moskalik, M., O'Regan, M., Pawłowska, J., Promińska, A., Rachold, V., Vonk, J. E., & Werner, K. (2015). Variability in transport of terrigenous material on the shelves and the deep Arctic Ocean during the Holocene. *Polar Research*, *34*(1). https://doi.org/10.3402/POLAR.V34.24964

Wheeler, P. A., Watkins, J. M., & Hansing, R. L. (1997). Nutrients, organic carbon and organic nitrogen in the upper water column of the Arctic Ocean: implications for the sources of dissolved organic carbon. *Deep Sea Research Part II: Topical Studies in Oceanography*, *44*(8), 1571–1592. https://doi.org/10.1016/S0967-0645(97)00051-9

Wikner, J., & Andersson, A. (2012). Increased freshwater discharge shifts the trophic balance in the coastal zone of the northern Baltic Sea. *Global Change Biology*, *18*(8), 2509–2519. https://doi.org/10.1111/J.1365-2486.2012.02718.X

Wild, B., Andersson, A., Bröder, L., Vonk, J., Hugelius, G., McClelland, J. W., Song, W., Raymond, P. A., & Gustafsson, Ö. (2019). Rivers across the Siberian Arctic unearth the patterns of carbon release from thawing permafrost. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(21), 10280–10285. https://doi.org/https://doi.org/10.1073/pnas.1811797116

Wold, H. (1975). Path Models with Latent Variables: The NIPALS Approach. *Quantitative Sociology*, 307–357. https://doi.org/10.1016/B978-0-12-103950-9.50017-4

Woo, M. K., Kane, D. L., Carey, S. K., & Yang, D. (2008). Progress in permafrost hydrology in the new millennium. In *Permafrost and Periglacial Processes* (Vol. 19, Issue 2, pp. 237–254). https://doi.org/10.1002/ppp.613

Wünsch, U. J., Geuer, J. K., Lechtenfeld, O. J., Koch, B. P., Murphy, K. R., & Stedmon, C. A. (2018). Quantifying the impact of solid-phase extraction on chromophoric dissolved organic matter composition. *Marine Chemistry*, *207*, 33–41. https://doi.org/10.1016/J.MARCHEM.2018.08.010

Wünsch, U. J., Murphy, K. R., & Stedmon, C. A. (2017). The One-Sample PARAFAC Approach Reveals Molecular Size Distributions of Fluorescent Components in Dissolved Organic Matter. *Environmental Science and Technology*, *51*(20), 11900–11908. https://doi.org/10.1021/acs.est.7b03260

Yan, G., & Kaiser, K. (2018a). Ultralow Sample Volume Cupric Sulfate Oxidation Method for the Analysis of Dissolved Lignin. *Analytical Chemistry*, *90*(15), 9289–9295. https://doi.org/10.1021/acs.analchem.8b01867

Yan, G., & Kaiser, K. (2018b). A rapid and sensitive method for the analysis of lignin phenols in environmental samples using ultra-high performance liquid chromatography-electrospray ionization-tandem mass spectrometry with multiple reaction monitoring. *Analytica Chimica Acta*, *1023*, 74–80. https://doi.org/10.1016/j.aca.2018.03.054

Yang, L., Hur, J., & Zhuang, W. (2015). Occurrence and behaviors of fluorescence EEM-PARAFAC components in drinking water and wastewater treatment systems and their applications: A

review. *Environmental Science and Pollution Research*, *22*(9), 6500–6510. https://doi.org/10.1007/S11356-015-4214-3

Zielinski, O., Rüssmeier, N., Ferdinand, O. D., Miranda, M. L., & Wollschläger, J. (2018). Assessing Fluorescent Organic Matter in Natural Waters: Towards In Situ Excitation–Emission Matrix Spectroscopy. *Applied Sciences 2018, Vol. 8, Page 2685*, *8*(12), 2685. https://doi.org/10.3390/APP8122685

Zimmermann, W. (1990). Degradation of lignin by bacteria. *Journal of Biotechnology*, *13*(2–3), 119–130. https://doi.org/10.1016/0168-1656(90)90098-V

Paper A

Terrestrial dissolved organic matter mobilized from eroding permafrost controls microbial community composition and growth in Arctic coastal zones

Anders Dalhoff Bruhn, Colin A. Stedmon, Jérôme Comte, Atsushi Matsuoka , Niek Jesse Speetjens, George Tanski , Jorien E. Vonk and Johanna Sjöstedt.

This paper has been published in:

Anders D. Bruhn, Colin A. Stedmon, Jeromé Comte, Atsushi Matsuoka, Niek J. Speetjens, George Tanski, Jorien E. Vonk and Johanna Sjöstedt (2021): Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones. Front. Earth Sci. 9:640580. doi: 10.3389/feart.2021.640580

frontiers in Earth Science

ORIGINAL RESEARCH
published: 24 March 2021
doi: 10.3389/feart.2021.640580

Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones

Anders Dalhoff Bruhn[1], Colin A. Stedmon[1], Jérôme Comte[2], Atsushi Matsuoka[3,4], Niek Jesse Speetjens[5], George Tanski[5,6], Jorien E. Vonk[5] and Johanna Sjöstedt[1,7]*

[1] National Institute of Aquatic Resources, Technical University of Denmark, Lyngby, Denmark, [2] Centre - Eau Terre Environnement, Institut National de la Recherche Scientifique, Québec, QC, Canada, [3] Takuvik Joint International Laboratory (CNRS-ULaval), Québec, QC, Canada, [4] Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, United States, [5] Department for Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [6] Permafrost Research Section, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany, [7] Department of Biology, Aquatic Ecology, Lund University, Lund, Sweden

OPEN ACCESS

Edited by:
Louise Farquharson, University of Alaska Fairbanks, United States

Reviewed by:
Joanne K. Heslop, German Research Centre for Geosciences, Helmholtz Centre Potsdam, Germany
Alexander Michaud, Bigelow Laboratory for Ocean Sciences, United States
David E. Graham, Oak Ridge National Laboratory (DOE), United States

**Correspondence:*
Anders Dalhoff Bruhn
adbj@aqua.dtu.dk

Specialty section:
This article was submitted to Biogeoscience, a section of the journal Frontiers in Earth Science

Received: 11 December 2020
Accepted: 24 February 2021
Published: 24 March 2021

Citation:
Bruhn AD, Stedmon CA, Comte J, Matsuoka A, Speetjens NJ, Tanski G, Vonk JE and Sjöstedt J (2021) Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones. Front. Earth Sci. 9:640580. doi: 10.3389/feart.2021.640580

Climate warming is accelerating erosion along permafrost-dominated Arctic coasts. This results in the additional supply of organic matter (OM) and nutrients into the coastal zone. In this study we investigate the impact of coastal erosion on the marine microbial community composition and growth rates in the coastal Beaufort Sea. Dissolved organic matter (DOM) derived from three representative glacial deposit types (fluvial, lacustrine, and moraine) along the Yukon coastal plain, Canada, were used as substrate to cultivate marine bacteria using a chemostat setup. Our results show that DOM composition (inferred from UV-Visible spectroscopy) and biodegradability (inferred from DOC concentration, bacterial production and respiration) significantly differ between the three glacial deposit types. DOM derived from fluvial and moraine types show clear terrestrial characteristics with low aromaticity (S_r: 0.63 ± 0.02 and $SUVA_{254}$: 1.65 ± 0.06 L mg C^{-1} m^{-1} & S_r: 0.68 ± 0.01 and $SUVA_{254}$: 1.17 ± 0.06 L mg C^{-1} m^{-1}, respectively) compared to the lacustrine soil type (S_r: 0.71 ± 0.02 and $SUVA_{254}$: 2.15 ± 0.05 L mg C^{-1} m^{-1}). The difference in composition of DOM leads to the development of three different microbial communities. Whereas *Alphaproteobacteria* dominate in fluvial and lacustrine deposit types (67 and 87% relative abundance, respectively), *Gammaproteobacteria* is the most abundant class for moraine deposit type (88% relative abundance). Bacterial growth efficiency (BGE) is 66% for DOM from moraine deposit type, while 13 and 28% for DOM from fluvial and lacustrine deposit types, respectively. The three microbial communities therefore differ strongly in their net effect on DOM utilization depending on the eroded landscape type. The high BGE value for moraine-derived DOM is probably caused by a larger proportion of labile colorless DOM. These results indicate that the substrate controls marine microbial community composition and activities in coastal waters. This suggests that

biogeochemical changes in the Arctic coastal zone will depend on the DOM character of adjacent deposit types, which determine the speed and extent of DOM mineralization and thereby carbon channeling into the microbial food web. We conclude that marine microbes strongly respond to the input of terrestrial DOM released by coastal erosion and that the landscape type differently influence marine microbes.

Keywords: climate change, terrestrial dissolved organic matter, Arctic coastal zone, marine microbial community, chemostat, glacial deposits, permafrost

INTRODUCTION

The permafrost region in the northern hemisphere covers approximately 22% of the land surface which is not covered by glaciers and ice (Obu et al., 2019). A seasonally unfrozen active layer which thaws every year during the warm season is situated on top of the permafrost layer. The permafrost layer itself is permanently frozen, per definition, for at least 2 years in a row (Pollard, 2018). Recent estimates project that 1,000 ± 150 Pg organic carbon (OC) are stored in the upper 3 m of the soils, plus 500 Pg C in deeper deposits such as yedoma and deltaic sediments (Hugelius et al., 2014; Schuur et al., 2015; McGuire et al., 2018). The amount of carbon stored in active and permafrost layers across the Northern Hemisphere is larger than the carbon storage of any other soil regions on Earth (e.g., temperate and tropical soils) and also surpasses that of the atmosphere (Jobbágy and Jackson, 2000; Strauss et al., 2017).

Due to climate change, permafrost is warming at a global scale (Biskaborn et al., 2019) and the increased loss in soilintegrity primes Arctic coastlines for erosion (Günther et al., 2015; Hoque and Pollard, 2016; Obu et al., 2016; Fritz et al., 2017; Couture et al., 2018; Jones et al., 2018). Coastal erosion isfurther promoted by the reduction in landfast sea ice, making the shores vulnerable to environmental forcing, such as bathing of coastal bluffs in warm seawater and increased wave heights during storms (Overeem et al., 2011; Fritz et al., 2017). Dissolvedorganic matter (DOM), which is mobilized from permafrost compartments in Cryosols (i.e., soil type in cold regions which is affected by permafrost and cryoturbation processes), is highly biodegradable and directly available for microbial utilizationupon permafrost thaw (Dou et al., 2008; Vonk et al., 2013;Abbott et al., 2014; Drake et al., 2015; Fritz et al., 2015; Spenceret al., 2015; Wologo et al., 2020). Increased release of DOM from degrading permafrost landscapes to the Arctic coastal zone (ACZ) can potentially lead to changes in OC processingby marine microbes (Colatriano et al., 2018). This can in turn influence marine primary production and higher trophiclevels by increases in regenerated nutrients (Sipler et al., 2017), competition for mineral nutrients (Thingstad et al., 2008) and decreased light penetration (Arrigo and Brown, 1996). Utilization of dissolved organic carbon (DOC) (Tanski et al., 2019) and its offshore transport to deeper oceans (Belicka et al., 2002) will also be affected by the change in microbial carbon processing. This will ultimately affect the overall carbon cycle and thereby impact the carbon budget and aquatic ecosystems in the ACZ. Due to the magnitude of permafrost thaw (Romanovsky et al., 2010; Smith et al., 2010) and increase in coastal erosion (Günther et al., 2015; Fritz et al., 2017; Jones et al., 2018) a better estimate of marine microbial carbon processing of DOC from coastal erosion and its effect on microbial communities in ACZ is needed.

In this study we examine three different permafrost landscape units, which are omnipresent in our study region (Yukon coastal plain, western Canadian Arctic). The landscape units differ from their (post-)glacial genesis and OC storage and include fluvial (FLU) deposits (53 kg OC m^{-3} in the top meter of soil);lacustrine (LAC) deposits (47 kg OC m^{-3}); and moraine (MOR) deposits (40 kg OC m^{-3}) (Couture et al., 2018).
Organic matter in FLU deposits comes from dispersed plant detritus mixed with mineral soil or peat layers that accumulated under stagnant water conditions in abandoned river channels (Rampton, 1982; Schirrmeister et al., 2011). Organic matter in LAC deposits was incorporated from the reworking and deposition of older material eroded from shore bluffs into lakes together with *in situ* production of fresh OM from aquatic plants and animals (Schirrmeister et al., 2011) along with talik formation (Wolter et al., 2017) and OC degradation in these lake sediments (Walter Anthony et al., 2018) due to permafrost thaw under the lakes. For the MOR deposit type, peat lenses and organic-rich silt is characteristic for the upper part of the deposits, where syngenetic formation and cryoturbation of permafrost has been the main driver behind the storage of OM (Rampton, 1982). This lead to high rates of accumulation and preservation of plant remains, which consists of less decomposed OM and high TOC (Schirrmeister et al., 2011). The organic matter incorporated into FLU and LAC has likely experienced Holocene decomposition since they were waterlogged during this period, whereas MOR likely been preserved since the Pleistocene and experienced very low decomposition (Strauss et al., 2017). During Holocene decomposition, labile OM, such as aliphatic and peptide-like compounds, in FLU and LAC deposits have been reworked by microbial activity, thereby leaving a higher amount of less labile compounds, such as aromatic compounds, when refrozen during the cooling of the middle Holocene (Stapel et al., 2017; Strauss et al., 2017; Heslop et al., 2019).

Terrestrial OM along permafrost coasts is released primarily as particulate organic carbon (POC) to the ACZ (Tanski et al., 2016). Although erosion of the coastline delivers relatively low amounts of DOC (annual flux of 55 Mg yr^{-1}) to the ACZ compared to Arctic rivers (34–38 Tg y^{-1} if the entire pan-Arctic watershed is considered) (Guo and Macdonald, 2006; Holmes et al., 2012; Tanski et al., 2016; Wild et al., 2019), the DOC

released by coastal erosion is highly biodegradable upon thaw (Vonk et al., 2013, 2015; Fritz et al., 2015; Spencer et al., 2015). In addition, upon exposure to seawater with high ionic strengths, ion exchange can release and further dissolve mineral-bound particulate OM or colloids, which can mobilize 19–50% of the bound OM (Kaiser and Guggenberger, 2000; Kawahigashi et al., 2006; Dou et al., 2008). The release of DOM from coastal erosion can therefore be assumed to have a large influence on microbes in coastal environments. Multiple studies have shown that the supply of permafrost-derived DOM can lead to rapid changes in marine microbial community composition and growth (Blanchet et al., 2017; Sipler et al., 2017; Müller et al., 2018) and it has recently been shown that the marine bacteria *Chloroflexi*, have the capacity to utilize terrestrial DOM (Colatriano et al., 2018). The effect of rapid mineralization of OC from permafrost thaw has been shown for freshwater bacterial communities, where microbes can degrade 34% of permafrost-derived DOC within 14 days of incubation (Vonk et al., 2013) or 47% within 28 days of incubation (Mann et al., 2015). However, there are knowledge gaps associated with the impact of DOM derived from coastal erosion of different glacial deposit types and how it affects marine microbial composition and carbon processing.

Here we study the bioavailability of DOM, released directly to marine microbes in the ACZ, through coastal erosion, usinga chemostat approach. This allows the development of a stable microbial community under conditions where the constant supply of substrate could reflect conditions in coastal waters influenced by coastal erosion. We investigate the biodegradabilityof DOM derived from three representative glacial deposit types (FLU, LAC, and MOR) to test if differences in DOM character caninduce differences in bacterial community composition (BCC) and bacterial growth efficiency (BGE). We also follow the changes in DOM character imparted by microbes using absorbance and fluorescence spectroscopy. We hypothesized that (i) the DOM provided by erosion of permafrost coasts is highly bioavailable for marine microbes and that (ii) different DOM characteristics among glacial deposit types induce different growth rates and the establishment of different bacterial communities.

MATERIALS AND METHODS

Study Area

The study area is located in the western Canadian Arctic on the Yukon coastal plain near Herschel Island - Qikiqtaruk (**Figure 1**). The Yukon coastal plain is about 282 km long and 10–30 km wide. Cliff heights are diverse, ranging from 2–3 m on the mainland across from Herschel Island to 60 m in the eastern part of the Yukon coastal plains (Couture et al., 2018). The mean annual air temperature is −1 °C, with the highest temperature during July month reaching 7.8 °C (1971–2000) (Environment Canada, 2016). The parent material of the Cryosols in the Yukon coastal plains originate from an ice-rich glacial margin and has been deposited by glacial transportation (ice and meltwater) of earth material and later on reworked by water during Holocene Thermal Maximum (higher than modern summer temperatures) (Fritz et al., 2012).

Sampling targeted fluvial, lacustrine, and moraine deposits since they represent the permafrost landscape on the Yukon coastal plain and differ the most in OC content and storage among all post-glacial landscape units (Couture et al., 2018). Fluvial sediments are poorly-drained floodplain deposits consisting of fine-grained sediments. The fluvial sediments used for this study were taken from the floodplains of the Babbage River Delta and may resemble other floodplain deposits in the study area. The fine-grained and ice-rich fluvial deposits are prone to erosion and can be mobilized with increasing river discharge, floods and deltaic or river bank erosion. Lacustrine sediments originate from thermokarst formation within moraine deposits (Krzic et al., 2010). These sediments accumulated with lake formation and have a fine-grained composition with peat layers present (Rampton, 1982). Lacustrine deposits are situated in flat and gently sloping terrain and have a poor surface drainage (Rampton, 1982). Moraine deposits are very common and cover approximately 50% of the formerly glaciated part of the Yukon coastal plain. Moraines are ice-rich and surfaces usually well drained due to a distinct topographic gradient with slopes of 5–25 ° (Rampton, 1982). The finer sediments in the upper layer of moraine soil were commonly washed away by meltwater fromthe glacier and moved to lacustrine and marine environments (Krzic et al., 2010).

Collection of Samples

Permafrost and active layer sediment samples as well as coastal seawater samples were collected in April and May 2019. The sediment samples were taken with a SIPRE corer at three sites with different glacial deposit types; (A) fluvial (FLU, fluvial deposits, 69.20686 N, 138.36730W), (B) lacustrine (LAC, lacustrine-wetland deposits, 69.33580 N, 138.8768W), and (C) moraine (MOR, moraine ridge-drained deposits, 69.46122N, 139.24230W). Subsamples were taken from the active layer (FLU: 9–24 cm, LAC: 4–14 cm, and MOR: 27–42 cm) and from the permafrost layer (FLU: 105–120 cm, LAC: 55–70 cm, and MOR: 90–105 cm) of the cores. All samples were stored frozen (18 °C) and in the dark.

Seawater samples were collected in the wider Mackenzie Bay in Herschel Basin through the sea ice using a Niskin water sampler (69.50978 N, 138.86278 W). A Conductivity, Temperature, and Depth (CTD) cast was taken to make sure seawater was collected from the main water body avoiding layers below the sea ice or close to the seabed. Seawater was stored in pre-leached 1 L Nalgene bottles, which were rinsed with the sample two to three times. Seawater was stored cold (4 °C) in the dark for 1 month prior to the experiment. The seawater had a salinity of 31 ppt and was later used as a bacterial inoculum.

Preparation of Soil Medium and Bacterial Culture

All the equipment used in the experiment was acid cleaned. Soil extracts were prepared from the six soil cores from FLU, MOR, and LAC glacial deposit types. For each glacial deposit type, extracts were prepared by dissolving soils from the active layer (wet weight; FLU: 343 g, LAC: 331 g, and MOR: 282 g)

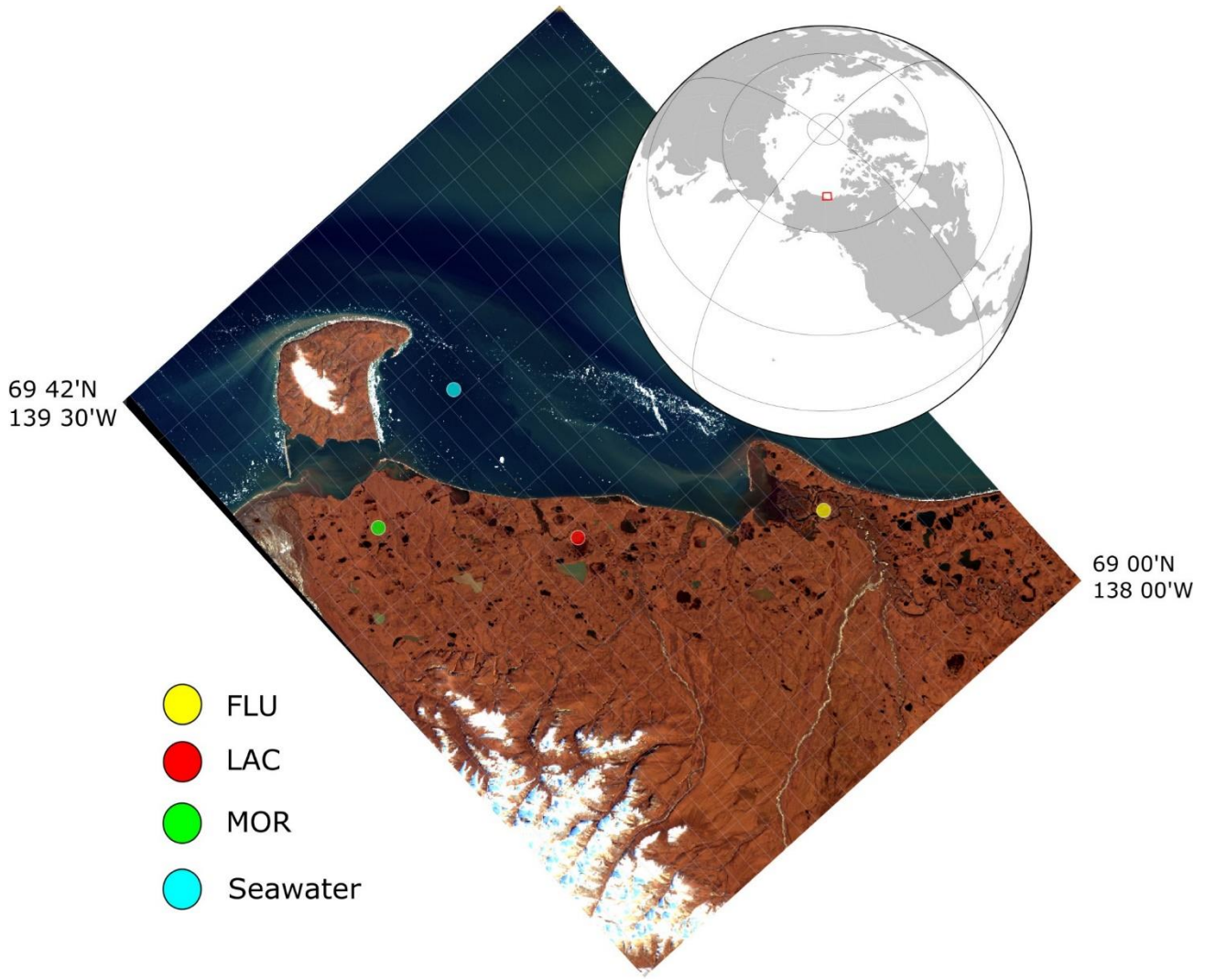


FIGURE 1 | Satellite image of sampling region in Northern Canada near Herschel Island – Qikiqtaruk along the Yukon Coastal Plain. Glacial deposit types (yellow = fluvial, red = lacustrine, and green = moraine) and coastal seawater (blue) were sampled in 2019. Background color: Landsat-8/OLI RGB image acquired on September 15, 2018. Contour interval: 3′.

or the permafrost layer (wet weight; FLU: 581 g, LAC: 192 g, and MOR: 222 g) separately into 8 L of ultrapure water (water extraction). The same volume of soil was added, which resulted in different mass due to differences in ice content between the glacial deposit types. The soil was kept in suspension overnight and then allowed to settle for 1 h before filtration througha 0.2 µm filter (AcroPak Capsules with Supor® Membrane, PALL). Salts were added (NaCl 21.1 g, $MgCl_2$ 4.5 g, Na_2SO_4 3.5 g, $CaCl_2$ 1 g, KCl 0.6 g, KBr 0.1 g, H_3BO_3 0.02 g, and $NaHCO_3$ 0.2 g per L) to a final salinity of 31 ppt (reflecting the same salinity and ratios as the sampled seawater). The water extractions from the soils were then stored at 20 ˚C overnight to let the salts completely dissolve. For each glacial deposit type, the soil extract from the active layer was mixed with an equal volume of soil extract from the permafrost layer (2 7 L mixed soil extracts per glacial deposit type) and then autoclaved. We mixed the two layers to reflect what will happen during coastal erosion when the whole soil column collapses into the ACZ. After the first autoclaving, the mixed soil extracts were splitinto four replicates (4 3.5 L mixed soil extracts per deposit type). The extracts were then re-filtered through a 0.2 µm filter and KNO_3 and K_3PO_4 was added to a final concentration of16 and 1 µM, respectively. The sterile soil extracts were then autoclaved a second time. The sterile soil extracts were stored overnight, in darkness at 4 ˚C, until the start of the experiment, where they were used as medium for the chemostat cultures.The mixed and sterile soil extracts are hereafter referred to as medium/media.

The sampled seawater obtained from Herschel Basin wasused as bacterial inoculum. To adapt the marine microbial communities to the DOM derived from the Cryosols and thereby get a faster establishment in the chemostats, 900 ml of seawater

was mixed with 200 ml from the respective autoclaved medium, 3 days before the start of each chemostat experiment. This formed the base of the culture to be used in the chemostat experiments.

Experimental Setup

Chemostats for each glacial deposit type were run in quadruplicates using the same approach as reported in Sjöstedt et al. (2012b). For each chemostat, 3.5 L of medium was used and 0.2 L of marine bacterial inoculum as culture. Mediumwas fed dropwise through a glass tube to prevent back growth (Hagstrom et al., 1984). Oxygen was supplied by passing air through a 0.2 μm pore-size polytetrafluoroethylene Acrodisc CR filter (Pall Corporation) (Zweifel et al., 1996). The chemostats were run for 14 days in the dark and at 20 ˚C, which is standard temperature for biodegradation experiments (Vonk et al., 2015). The dilution rate was 1 day^{-1}, which is close to the median growth rate of marine bacteria ranging from polar to temperate regions (Moriarty, 1986).

Samples for bacterial abundance (BA) and optical measurements of DOM (absorbance and fluorescence) were taken both from the culture (once per day) and the medium (every 2 days). Bacterial community composition samples were taken from the culture every day. Samples for DOC and nutrient concentrations were taken from the culture daily and from the medium at day 0, 7 and 14. Samples from the culture were taken from the outflow of the chemostat, whereas samples from the medium were taken by carefully disconnecting the tubing feeding the culture with medium.

DOM Absorbance and Fluorescence

Samples (20 ml) were collected in acid washed and precombusted 40 ml glass vials with Teflon-lined caps and measured within 2 h of collection. The absorption spectrum of colored DOM (CDOM) and excitation-emission-matrix (EEM) of fluorescent DOM (FDOM) were measured on filtered samples (0.2 μmSupor Acrodisc, Pall Corporation). Both CDOM and FDOMwas measured on an Aqualog fluorometer (HORIBA Jobin Yvon). The CDOM absorbance was measured in a 1 cm quartz cuvette between 239 and 800 nm, with increment of3 nm and integration time of 0.1 s. The FDOM fluorescence was measured in the same 1 cm quartz cuvette immediatelyafter the CDOM measurement. Emission wavelengths were 245–824 nm with increment of 2 nm and excitation wavelengths 240–450 nm with increment of 5 nm. Theemission integration time was adjusted to account for varying FDOM fluorescence intensities between samples. An ultrapure water sample was used as a blank for both absorbance andfluorescence measurements.

The absorption spectrum of the ultrapure blank was subtracted from the absorption spectrum of the chemostat samples. This was performed by subtracting the absorption meanvalue between 590 and 800 nm. The absorption values were converted to the naperian absorption coefficient according to the following equation, $a_\lambda = 2.303^{*}A/L$, where A (unitless)is the optical density measured by the instrument and L (m)is the length of the cuvette and λ is wavelength. Fluorescence data was converted into Raman units (R.U) applying Raman calibration (Lawaetz and Stedmon, 2009), blank subtracted and then corrected for inner filter effects using the drEEM toolbox (Murphy et al., 2013). The quality of CDOM in the contrasting media were compared by using the spectral slope coefficients (Helms et al., 2008) and the carbon specific UV absorbance at 254 nm, $SUVA_{254}$ (based on decadic absorbance) (Weishaar et al., 2003). The intensity of CDOM absorption in the media and the cultures was compared using the absorption at 330 nm (highest observed degradation wavelength). The fluorescence intensities were summarized using the peak regions defined by Coble (1996). An attempt to characterize the fluorescence data using parallel factor analysis was unsuccessful. A robust result could not be obtained largely due to the fact that the changes in fluorescence occurring were very minor and that the sample dataset largely represented replicate samples of same medium, which invalidated the use of the commonly applied split half analysis for testing the result.

The steady state period for the cultures were determined based on achievement of a stable CDOM absorbance signal(330 nm) after initial flushing and establishment of thecultures during the first week (between days 7 and 13, **Supplementary Figure 1**).

DOC and Nutrients

After filtration, subsamples for DOC, total dissolved nitrogen (TDN), nitrate/nitrite and phosphate were collected in 30 mL acid-washed HDPE bottles and stored at −20 ˚C until analysis. DOC and TDN concentrations were determined by high-temperature combustion (720 ˚C) using a Shimadzu TOC-V CPH-TN carbon and nitrogen analyzer. The instrument was calibrated using acetanilide (Cauwet, 1999) and carbon determination was referenced to the community deep-sea reference (Hansell laboratory, Miami). DOC accuracy was assured by comparing to reference material provided by the Hansell Laboratory (Sargasso Sea water from 2,600 m near Bermuda; DOC concentration of 42–44 μM). Precision was estimated using a (more comparable) higher concentration sample of aged seawater. This was analyzed every 8 samplesin the runs, and was determined to be 7 μM. Nitrite, nitrate and phosphate were analyzed on a SmartChem200 discrete analyzer (AMS Alliance). The combined concentration of nitrate and nitrite was determined using the method described in Schnetger and Lehners (2014). Phosphate was measured using the manual method described in Hansen and Koroleff (1999). Nutrients standard kits from OSIL were used for calibration.

Bacterial Abundance

Samples (1.6 ml) for BA were fixed with glutaraldehyde (1% final conc.) and stored at 20 ˚C until measured in the flow cytometer. Samples were stained with SybrGreen (Invitrogen) and cells were counted on a FACSCantoII flow cytometer (BD Biosciences), as previously described (Gasol and Del Giorgio, 2000). Fluorescent beads (BD Trucount Tubes, BD Biosciences) were used to calibrate the flow rate. The data was processed usingFlowJo9.9.5. Average bacterial abundances were calculated for the steady state period (days 7–13). Bacterial carbon production

(BCP) was calculated based on the bacterial abundance as cells produced per day multiplied by a carbon constant ($1.6^{*}10^{-15}$ moles carbon $cell^{-1}$ (Lee and Fuhrman, 1987).

Bacterial Community Composition

Samples for BCC analysis were taken from the cultures from each deposit type every day. 100 ml of water was filtered through 0.2 µm Supor filters (25 mm; Pall Corporation) to collect bacterial biomass for DNA extraction. The filters were stored at−20 ˚C until extraction. The samples from days 0, 9, 11, and 13 for each culture was selected for analysis. From these filters, DNA was extracted using the Qiagen Power Soil kit.

Amplification of the 16S rRNA gene, equimolar pooling and sequencing was performed at the Plateforme d'analyses génomiques (IBIS, Université Laval, Quebec City, QC, Canada). Amplification of the 16S V3–V4 regions was performed using the sequence specific regions (341F-805R) described in Herlemann et al. (2011) using a two-step dual-index PCR approach specifically designed for Illumina instruments. In the first step, the gene specific sequence is fused to the Illumina TruSeq sequencing primers and PCR was carried out in a total volume of 25 µL that contains 1×Q5 buffer (NEB), 0.25 µM of each primer, 200 µM of each dNTPs, 1 U of Q5 High-Fidelity DNA polymerase (NEB) and 1 µL of template DNA. The PCR started with an initial denaturation at 98 ˚C for 30 s followed by 35 cycles of denaturation at 98 ˚C for 10 s, annealing at 55 ˚Cfor 10 s, extension at 72 ˚C for 30 s and a final extensionat 72 ˚C for 2 min. The PCR reaction was purified using the Axygen PCR cleanup kit (Axygen). Quality of the purified PCR products was checked on a 1% agarose gel. A 50–100 folddilution of this purified product was used as a template fora second PCR step with the goal of adding barcodes (dual-indexed) and missing sequence required for Illumina sequencing. Cycling for the second PCR were identical to the first PCRbut with 12 cycles. PCR reactions were purified as above, checked for quality on a DNA7500 Bioanlayzer chip (Agilent) and then quantified spectrophotometrically with the Nanodrop 1000 (Thermo Fisher Scientific). Barcoded amplicons were pooled in equimolar concentration for sequencing on the Illumina Miseq.

Quality control of the sequences was performed by first removing the primer sequence using the Cutadapt 2.7 tool (Martin, 2011) and secondly by trimming the sequences at 260 bp (forward sequence) and 190 bp (reverse sequence). Sequence analysis was then performed using the IBIS computing infrastructure and the Dada2 algorithm (Callahan et al., 2016). In the end, 1,694,984 sequences were retained (median of 42,939 sequences per sample) following the different quality filters resulting in a total of 912 amplicon sequence variants (ASV). Taxonomic assignment was performed using the SILVA database (v.138) (Quast et al., 2013). DNA sequences have been deposited in the National Centre for Biotechnology Information (NCBI) Sequence Read Archive under accession number PRJNA675030.

A statistical method applied for differential gene expressions was used to rank ASVs consistently present in replicates as representative for each deposit type (Robinson and Smyth, 2008). In brief, a tagwise dispersion function [edgeR, R package (Robinson et al., 2010)] was used to rank ASVs according to their consistency among replicates and analyze which ASVs that differed significantly between deposit types. Differences were considered significant at $p < 0.01$. By using a generalized linear model, we tested for differential representation of ASVs between deposit types using the Toptag function, an analysis quite similar to an ANOVA. The analysis applies log_2-counts per million (logCPM) that is used for estimating relative representation in the community, where low value within a range from 1 to 100is considered high relative abundance. The analysis also reports logFC that is the *x*-fold change in ASV contribution to the community. The change in log_2 CPM gives a measure of the consistency of replicates (Canelhas et al., 2016).

Richness (S.Obs, observed number of species) and evenness were calculated in R 3.0.2 using the package Vegan (Oksanen et al., 2019). Non-metric multidimensional scaling (NMDS) was performed using normalized data (relative abundances) of ASVs obtained from sequencing of the 16S rRNA gene.Distances were based on Bray–Curtis dissimilarity matrix to describe relationships in community composition among samples. Vectors representing significantly correlated (PERMANOVA, p 0.05, df = 39, n = 999 permutations)DOM characteristics were plotted with the ordination [R 3.0.2, Vegan package (Oksanen et al., 2019)]. Vector lengthand direction reflects strength and direction of correlation between the DOM characteristics and community composition. Correlations between differences in community composition and differences in DOM utilization (BCP, BGE, and changein DOM characteristics between medium and culture) was investigated using Mantel tests [R 3.0.2, mvabund package, (Wang et al., 2012)].

Biological Oxygen Demand (BOD)

To investigate respiration rates, a sample from each culture replicate in each glacial deposit type (n = 12) was taken atthe end of the experiment and incubated for 24 h. Oxygen consumption was measured using a Pyroscience Firesting four channel optode oxygen sensor equipped with 20 mL respiration vials and a temperature sensor. As the chemostats for each glacial deposit type were set up sequentially (1 day apart), the respiration measurements could all be carried out on day 13 (steady state period). Medium and culture from each glacial deposit typewas mixed 1:1, to ensure adequate substrate supply during the respiration measurements. Respiration rates were calculated as the slope of oxygen concentration over time for two periods (2–12 h and 12–24 h) as for several of the incubations it was clear that there was a shift (flattening) of the oxygen concentrations during the BOD experiment. Data from the first 2 h were disregarded as it was clear there was a lag time before the establishment of stable respiration rates.

Data Analysis, Presentation, and Statistics

To simplify the data analysis only measurements from the steady state period were used to compare the growth and uptake rates of the cultures. For the optical measurements (absorption and fluorescence), the absorption spectrum and the EEM from the

culture were subtracted with the absorption spectrum and the EEM from their respective medium for each day in the steady state period. The newly created absorption spectra and EEMs from the subtraction could then be used to identify the net production or removal of CDOM and FDOM. A mean value per deposit type was calculated using replicates and days in the steady state.

To investigate the fate of bioavailable carbon from the glacial deposit types, four parameters were calculated; (1) total respiration (calculated from BOD experiments as the sum of respiration rates <12 h and >12 h multiplied by 12 h), (2) total bacterial carbon production [based on bacterial abundance as cells produced per day multiplied by a carbon constant of $1.6^{*}10^{-15}$ moles carbon $cell^{-1}$ (Lee and Fuhrman, 1987)], (3) the total change in DOC concentration (calculated from the difference between medium and culture) and (4) BGE (calculated by dividing BCP by the DOC uptake).

In order to test for differences between glacial deposit types an analysis of variance (ANOVA) test was performed when criteria of normality and heterogeneity was met. If these criteria were not met, the Kruskal–Wallis test was applied. To resolve regions of the absorption spectra and fluorescence EEMs where medium and culture differed significantly, a two sample *t*-test was applied for each wavelength independently. The significance level was set to a *p*-value of 0.05. ANOVA and *t*-test ran in MatLab while Kruskal–Wallis test was applied in R.

The averages (mean) and standard deviations were calculated across the four replicates within each glacial deposit type. This was applied to all parameters using all measurements in the steady state (across days and replicates) unless else stated. Difference in samples number are due to missing data and outliers (defined from boxplots in R) and sample numbers willbe stated after each parameter for each glacial deposit type.

RESULTS

Characteristics of DOM

The DOC concentration in the media was highest in LAC, followed by MOR and FLU with an average concentration of 632 (34, n = 8), 543 (36, n = 8), and 526 (26, n = 6) µM DOC, respectively (**Figure 2**). DOC concentrations in the LAC medium were significantly higher than in both the MOR and FLU medium (ANOVA, $p < 0.01$), while MOR and FLU were not significantly different (ANOVA, $p = 0.16$). One replicate of FLU (2 out of 8 measurements) was removed when calculating the mean DOC concentration due to its anomalously low value (368 µM), while the source of the error could not be identified. The average DON concentration in the media was 28.7 (4.5, n = 8), 25.2 (1.9, n = 8), and 24.1 (2.7, n = 8) µM DON for FLU, LAC, and MOR, respectively (**Figure 2**) and there was no significant difference among the three media (ANOVA, $p > 0.05$). The DOC:DON ratios for FLU, LAC, and MOR media were 17.0 (±1.2, n = 8), 25.1 (±1.4, n = 8), and 22.8 (±2.7, n = 8), respectively. The DOC:DON ratio for FLU was found to be significantly lower than those of LAC and MOR (ANOVA, $p < 0.01$), while LAC and MOR did not differ significantly between each other (ANOVA, $p > 0.05$).

The LAC medium had the highest absorption coefficients (**Figure 3**). The absorption coefficients values at 330 nm differed significantly between all three media and were 15.5 m^{-1} (±0.4, n = 8) for LAC, 9.2 m^{-1} (±0.5, n = 12) for FLU, and 7.3 m^{-1} (±0.5, n = 11) for MOR (**Figure 3**, ANOVA, $p < 0.01$). Spectral slopes (S) and $SUVA_{254}$ were used to assess the quality of the CDOM (**Table 1**). The S values for 275–295 nm (UVB) and 350–400 nm (UVA) ranged between 13.4–13.7 and 19.2–21.3 μm^{-1} in the three media. The spectral slopes for the UVB area did not differ significantly between the three media (ANOVA, $p > 0.05$), while the spectral slope of the UVA area was significantly higher in FLU medium than in the other two media (ANOVA, and $p < 0.05$). The spectral slope ratio (S_r) values differed significantly between all media (ANOVA, $p < 0.01$) and were highest for LAC followed by MOR and FLU. The $SUVA_{254}$ values (**Table 1**) were also significantly different between all media (ANOVA, $p < 0.01$) and were highest for LAC, followed by FLU and MOR.

The fluorescence characteristics of the media are shown in **Figure 3** and the fluorescence intensities for common peak regions (Coble, 1996) are reported in **Table 1**. As with the absorption spectra, the fluorescence intensities differed between the media. The visible wavelength peaks (peak A, C, and M) were significantly different between all media (ANOVA, $p < 0.01$) with highest values for LAC, followed by FLU and MOR (**Table 1**). For peak T, the fluorescence intensities did not differ between FLU and LAC (ANOVA, $p > 0.05$), while it was significantly lower for MOR compared to the other two (ANOVA, $p < 0.01$) (**Table 1**). For peak B the fluorescence intensities did not differ significantly between MOR and LAC (ANOVA, $p > 0.05$), while FLU had significantly higher intensity than the other two (ANOVA, $p < 0.01$) (**Table 1**).

Alteration of DOM in the Cultures

Inorganic nutrients were added to ensure replete conditions and carbon limitation in the chemostats. Measured nutrient concentrations in the media were on average 13–15 µM for combined nitrate and nitrite and 0.68–1.01 for phosphate, and remained above 6 and 0.3 µM, respectively in the water flowing out of the cultures.

There were a significant difference in the concentrations of DOC and DON between the medium and the culture for all glacial deposit types (ANOVA, $p < 0.05$). The total removal of DOC in FLU, LAC, and MOR cultures was 132 (50, n = 28), 160 (46, n = 28), and 158 (48, n = 28) µM (**Table 2**) corresponding to 24% (9%), 25% (7%), and 29% (8%) of the initial DOC, respectively. For DON, the highest removal was instead observed for FLU culture and was found to be 10 (4, n = 28) µM corresponding to 36% (8%) loss, whereas for LACand MOR cultures it was found to be 4 (2, n = 28) and 4(1, n = 28) µM corresponding to a loss of 17% (6%) and 18% (3%) respectively. The DOC:DON ratios in the cultures changed from their corresponding media and were calculatedto be 22 (4.5, n = 28), 23 (1.9, n = 28) and 20.8 (6.2, n = 28) for respectively FLU, LAC, and MOR cultures, however, the change was only found to be significant for the FLU deposit type (ANOVA, $p < 0.01$).

The change in the absorption spectra (**Figure 4**) reveals the spectral fingerprint on the CDOM imparted by the

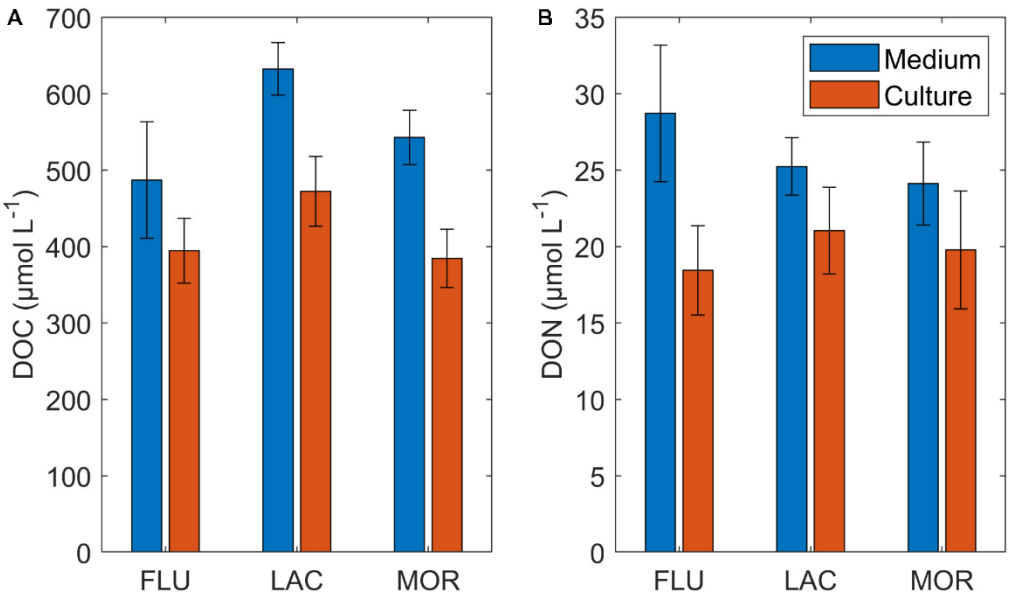


FIGURE 2 | Average DOC (A) and DON (B) Concentration (µmol L^{-1}) of soil media and cultures at steady state. NO_3^- and NO_2 are subtracted from DON, while ammonium is still included. Error bars represent the standard deviation.

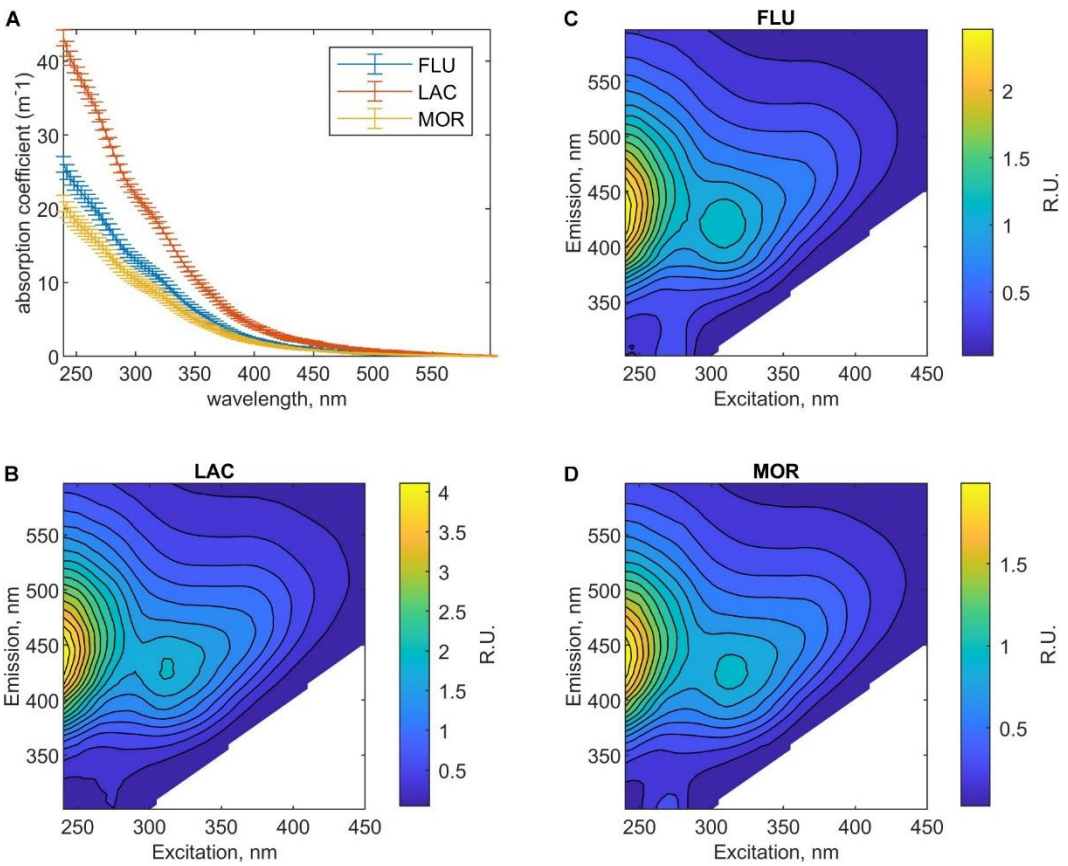


FIGURE 3 | Mean absorption spectra (A) and fluorescence EEMs (B–D, displayed as Raman units, R.U.) for the filtered media at steady-state. (B) LAC, (C) FLU, and (D) MOR.

TABLE 1 | Mean spectral slopes [S(275–295) and S(350–400)], slope ratios (S_r), $SUVA_{254}$ values (calculated from the soil medium CDOM spectra) and intensities of five commonly reported fluorescence peaks (Coble, 1996) (peak *A* = Ex/Em: 260/430, peak *B* = Ex/Em: 275/305, peak *C* = Ex/Em: 340/440, peak *M*: Ex/Em: 300/390, and peak *T* = Ex/Em: 275/340) for the three glacial deposit types.

DOM characteristics	Proxy	FLU	LAC	MOR
S(275–295) (**μ**m^{-1})	Molecular weight; photobleaching	13.4 (±0.29)	13.7 (±0.22)	13.4 (±0.31)
S(350–400) (**μ**m^{-1})	Molecular weight; photobleaching	21.3 (±0.69)	19.2 (±0.59)	19.6 (±0.70)
S_r	Molecular weight; origin of water	0.63 (±0.02)	0.71 (±0.02)	0.68 (±0.00)
$SUVA_{254}$ (L mg C^{-1} m^{-1})	Aromaticity	1.65 (±0.06)	2.15 (±0.05)	1.17 (±0.06)
Peak A (R.U.)	Humic-like compounds; terrestrial	1.73 (±0.05)	2.79 (±0.05)	1.38 (±0.08)
Peak B (R.U.)	Tyrosine and protein-like compounds; autochthonous	0.50 (±0.04)	0.29 (±0.02)	0.32 (±0.15)
Peak C (R.U.)	Humic-like compounds; terrestrial	0.95 (±0.03)	1.69 (±0.04)	0.85 (±0.06)
Peak M (R.U.)	Marine humic-like compounds;	1.07 (±0.04)	1.42 (±0.03)	0.76 (±0.05)
Peak T (R.U.)	Tryptophan and protein-like compounds; autochthonous	0.40 (±0.03)	0.39 (±0.01)	0.24 (±0.03)

R.U. corresponds to Raman units. Numbers in bold represent mean values, while numbers in parentheses represent standard deviations.

TABLE 2 | Overview of the carbon processing at 24 h timescale due to microbial activity in the cultures.

Glacial deposit type	Respiration (**μ**M O_2 d^{-1})	Bacterial carbon production (**μ**M C d^{-1})	DOC uptake (**μ**M C d^{-1})	BGE (%)
FLU	90 (±12, n = 8)	16 (±13, n = 26)	132(±50, n = 28)	13(±12, n = 25)
LAC	57 (±20, n = 8)	48 (±37, n = 26)	160 (±46, n = 28)	28(±20, n = 23)
MOR	111 (±17, n = 8)	105 (±58, n = 28)	158 (±48, n = 28)	66(±42, n = 26)

The numbers are calculated as the mean value over the whole steady state period. Numbers in bold represent mean values, while numbers in parentheses represent standard deviations. The standard deviation indicates the variation across all the samples measured from different days and replicates.

marine microbial communities in the cultures, with a negative change corresponding to degradation and a positive change to production of CDOM. For FLU, the absorption between 239 and 450 nm was significantly lower in the culture compared to the medium during the steady state period (t-test, $p < 0.05$). A similar yet weaker change was apparent for LAC and only significant at wavelengths between 308 and 371 nm (t-test, $p < 0.05$). The reduction in CDOM absorption at 330 nm, for FLU and LAC was respectively 1.32 (±0.28, n = 23) m^{-1} and 0.70 (±0.35, n = 22) m^{-1} and significantly different (ANOVA, $p < 0.01$). For MOR, there was a significant production in CDOM absorption between 239 and 296 nm (t-test, $p < 0.05$, n = 24) and no significant change at 330 nm (t-test, $p > 0.05$).

The fluorescence properties of DOM were altered by the marine microbial communities, but the extent to which the EEM area was altered differed in between all three glacial deposit types (**Figures 5A–C**). The change in fluorescence intensity between FLU medium and culture was significant for all peaks (t-test, $p < 0.05$) and the change in peak A, peak B, peak C, peak M, and peak T, corresponded, respectively, to −0.15 (± 0.07, n = 23), −0.18 (± 0.04, n = 23), − 0.06 (± 0.04, n = 23), − 0.14 (±0.04, = 23), and −0.05 (±0.05, n = 23) R.U. (**Figure 5A**). For LAC, the change in fluorescence intensity was only significant for peak B and peak M (t-test, $p < 0.05$) and corresponded, respectively, to 0.05 (± 0.02, n = 22) and − 0.07 (± 0.04, n = 22) R.U. (**Figure 5B**). As seen from **Figures 5D,G**, both the spectral characteristics of the fluorescence loss in the peak A and M regions in the FLU and LAC deposit types was very similar, although the microbial community in FLU removed much more from the medium (**Table 1**). The spectral characteristics of the fluorescence changein peak B region was also similar between FLU and LAC cultures (**Figure 5E**), however, there was a large removal in the FLU culture, while there was a weak, but still significant production in the LAC culture. The fluorescence loss in the region of peak C was comparable in character and intensity of removal for both FLU and LAC cultures (**Figure 5F**), even though there were differing start concentrations in the respective media (**Table 1**). For MOR, we observed no significant removal of fluorescence but rather

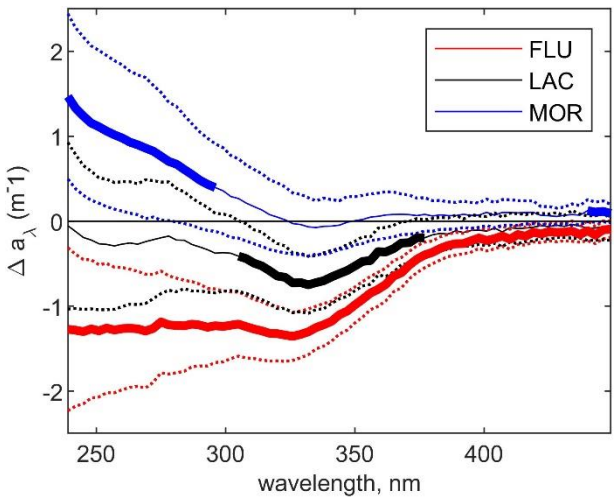


FIGURE 4 | Mean absorbance change between culture and medium at steady-state for each glacial deposit type. The dotted lines are displaying the standard deviation across all measurements in the steady-state period for each glacial deposit type (n = 28). The bold line is showing the area of the absorption spectrum that has significantly changed (t-test, $p < 0.05$).

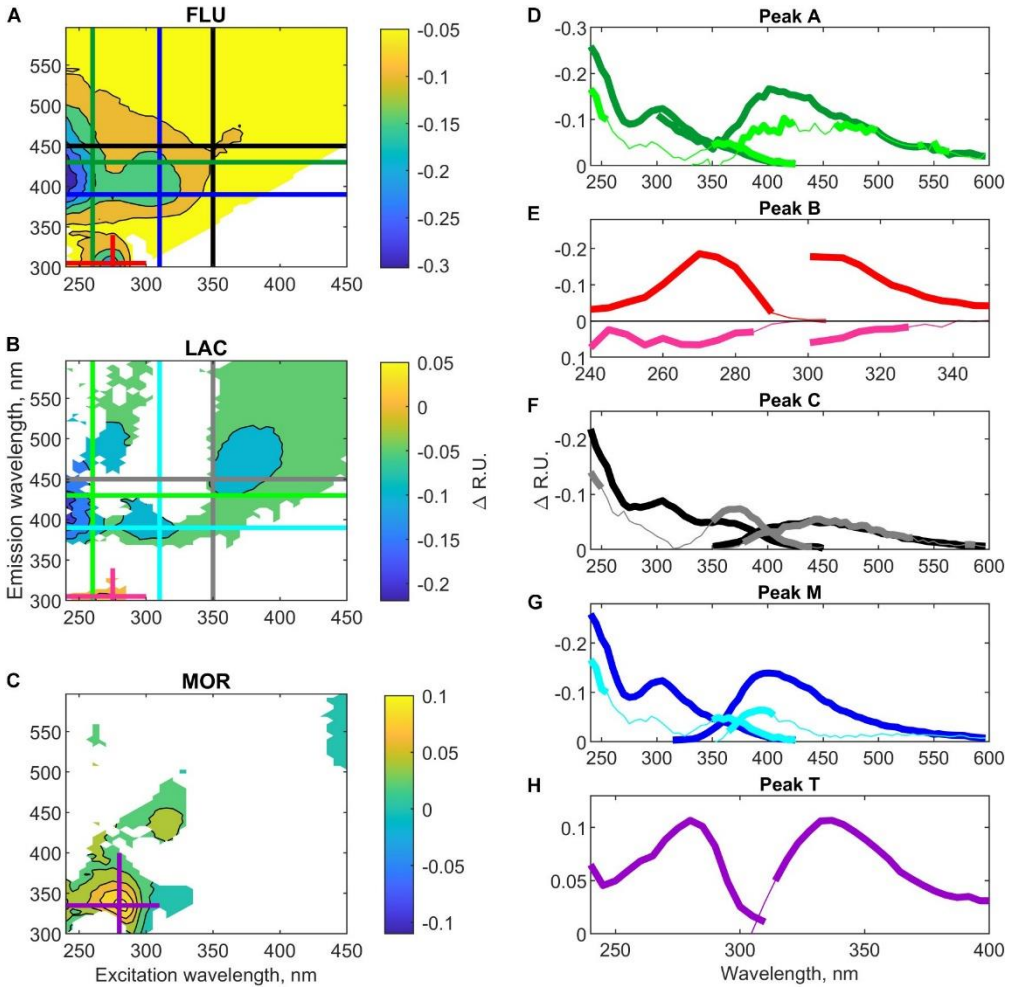


FIGURE 5 | Mean fluorescence change (Raman units, R.U.) between culture and medium at steady-state for each glacial deposit type. Panels (A–C) Shows the excitation-emission-matrix (EEM) with areas that are significant ($p < 0.05$) changed from the medium. (A) FLU, (B) LAC, and (C) MOR. Panels (D–H) Shows the excitation and emission spectra at specific emission and excitation wavelengths for different peaks. Panel (D) Peak A, (E) Peak B, (F) Peak C, (G) Peak M, and (H) peak T. The bold line shows which area is significant different from the soil medium. Note the *y*-axis changes, negative values represent a removal of fluorescence and positive values a production.

a significant production of peak T fluorescence by 0.10 (±0.03, $n = 24$) R.U. (**Figures 5C,H**).

Bacterial Growth and Bacterial Community Composition

At steady state the average BA was significantly different between all the three glacial deposit types (Kruskal–Wallis, $p < 0.005$) and the average abundance was 9.2 (±2.5, $n = 28$) 10^6 cell/mL in FLU culture, 32.1 (±14.6, $n = 28$) 10^6 cell/mL in LAC culture and 45.7(1±4.2, $n = 28$) 10^6 cell/mL in MOR culture.

Analysis of BCC was based on the % of reads. The BCC in the start cultures (**Figure 6A**) were dominatedby *Gamma-* and *Alphaproteobacteria* in all cultures. In the FLU start culture, the bacterial community was composedof 48% *Gammaproteobacteria*, 42% *Alphaproteobacteria* and 10% *Bacteroidetes*. For LAC, the start culture had a BCC of 91% *Gammaproteobacteria*, 8% *Alphaproteobacteria* and 1% *Bacteriodetes*. Last, the MOR start culture had a BCC of 61% *Gammaproteobacteria*, 36% *Alphaproteobacteria,* and 3% *Bacteroidetes*. Other taxa contributed to less than 0.5% of the sequences in each of the three glacial deposit types. In the start cultures, the dominant order within *Alphaproteobacteria* was *Rhodobacterales* which contributed to 99, 87, and 98% of the *Alphaproteobacterial* reads in FLU, LAC, and MOR, respectively. In LAC start culture, the abundance of *Sphingomonadales* was 11% (**Supplementary Table 1**). Within *Gammaproteobacteria* the dominant orders in all treatments were *Oceanospirillales* and *Burkholderiales*. In the MOR start culture, *Cellvibrionales* and *Nitrosococcales* were also important and the abundance of each of these orders were 15% (**Supplementary Table 1**).

During the steady state period (**Figure 6B**), the bacterial community in the FLU and LAC cultures became dominated by *Alphaproteobacteria* [67% (±5, $n = 12$) and 87% (±13,

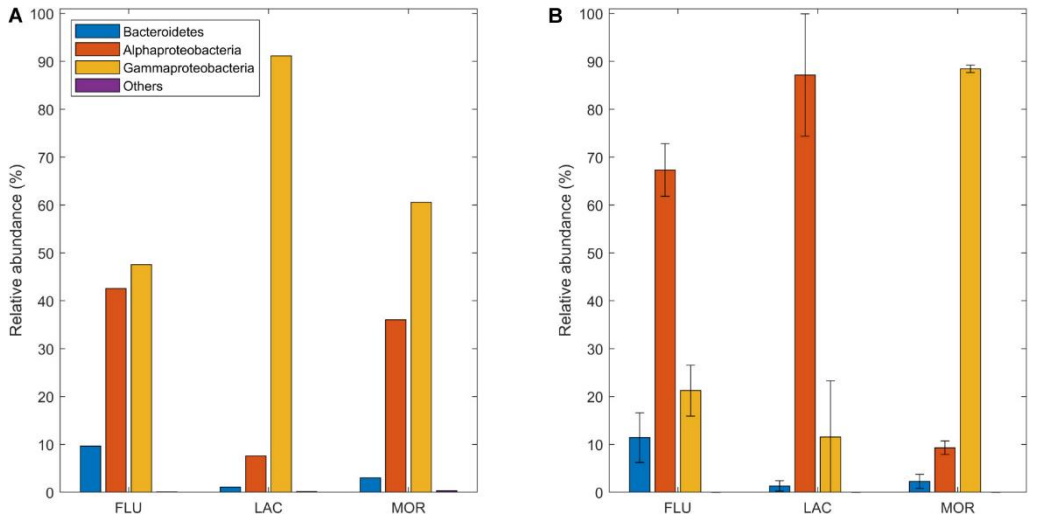


FIGURE 6 | Relative abundance of taxa (% of total number of reads) found in the culture for each glacial deposit type (A) Prior to the chemostat experiment and (B) During steady state of the chemostat experiment.

n = 12), for FLU and LAC, respectively]. In contrast the community in the MOR culture became more dominated by *Gammaproteobacteria* [88% (±, n = 12)]. The abundance of other taxa remained very low in the cultures from all three glacial deposit types. At steady state, *Rhodobacterales* was the dominant order within *Alphaproteobacteria* within all glacial deposit types and contributed to between 79 and 98% of the Alphaproteobacterial reads. Other orders contributing to at least 1% of the reads in at least one of the glacial deposit types were *Sphingomonadales*, *Caulobacterales* and *Rhizobiales* (**Supplementary Table 1**). Within *Gammaproteobacteria*, *Oceanospirillales* was the dominant order in the cultures from all glacial deposit types and contributed to between 79 and 99% of the Gammaproteobacterial reads. In the FLU and LAC cultures the abundance of *Nitrosococcales* was 14% and 13% of the reads, respectively. Other orders with an abundance above 1% in at least one of the glacial deposit types were *Alteromonadales* and *Cellvibrionales* (**Supplementary Table 1**).

Based on tagwise dispersion analysis, 125 ASVs differed significantly in abundance between the MOR and the FLU cultures, 157 ASVs between the LAC and FLU cultures and 144 ASVs between the LAC and MOR cultures (**Supplementary Tables 2–4**). Based on the top 10 ASVs there was a significant higher abundance of ASVs belonging to the genus *Sulfitobacter* in the FLU cultures compared to both the LAC and MOR cultures (**Supplementary Tables 2**, **3**). In the LAC culture several of the ASVs with significantly higher abundance compared to the FLU and MOR cultures belonged to the genus *Pacificibacter* and *Polaribacter* (**Supplementary Tables 3**, **4**). Whereas in the MOR culture several of the ASVs belonged to the genus *Amphritea* and *Marinomonas* (**Supplementary Tables 2**, **4**).

During the development of all the cultures, bacterial community richness decreased, while community evenness either increased slightly (FLU and LAC) or decreased drastically (MOR) (**Supplementary Table 5**). During the steady state, both richness and evenness were significantly higher in FLU compared to both LAC and MOR (Kruskal–Wallis, $p < 0.05$).

During the steady state period the cultures from the different glacial deposit types formed three clearly distinct clusters (**Figure 7**) indicating that the composition of bacterial communities in cultures was significantly different between glacial deposit types (PERMANOVA, F = 36.75, $r2$ = 0.67,$p < 0.01$), despite starting with the same marine inoculum. Replicates from the same glacial deposit type were very similar indicating good reproducibility in the BCC. For MOR and LAC cultures, there was a clear development from the start community to the steady state community, whereas for the FLU culture the start and steady state communities were rather similar. Significant DOM characteristics plotted as vectors on the NMDS plot indicate that BCC is linked to DOM composition of the medium (**Figure 7B** and **Supplementary Table 6**) and that the different glacial deposit types were linked to different DOM characteristics. The strongest predictors are peak A and peak C (R^2 0.9521, respectively, 0.9503). The BCC in the LAC culture was positively correlated to peak A and peak C but also to peak M and $SUVA_{254}$. The BCC in the FLU culture was positively correlated to peak B and negatively correlated to DOC concentration and S_r. In contrast, the BCC in the MOR culture was not positively correlated to any of the included DOM characteristics but negatively correlated to peak T. Bacterial community composition was also correlated to DOM utilization. Differences in BCC was significant correlated to BCP, BGE, ΔpeakA, ΔpeakB, ΔpeakC, ΔpeakM, and ΔpeakT (where Δ refer to the change between medium and culture, Mantel test, $p < 0.05$, Pearson r = 11–46%, n = 36, **Supplementary Table 7**).

Carbon Processing

Average respiration rates were calculated for two time periods, <12 and >12 h, as there was a significant (ANOVA, $p < 0.01$) decrease in rates for FLU and LAC (**Figure 8**). Initial respiration

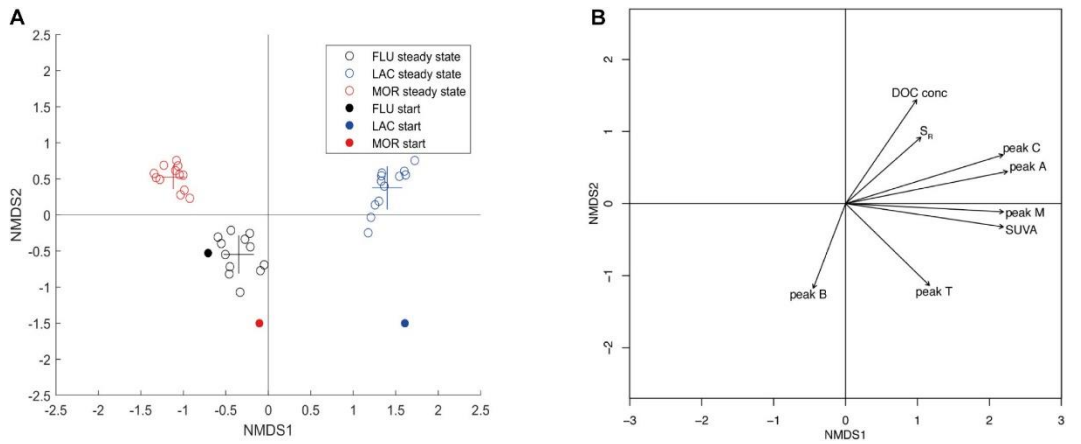


FIGURE 7 | (A) NMDS coordination plot based on 16S rRNA gene amplicon sequencing showing differences in bacterial community composition over time and between deposit types (Stress = 0.0903). Analysis is based on data from Illumina sequencing and Bray–Curtis dissimilarities. Glacial deposit type is indicated by color and time by filled and open circles (start and steady state). The center of the crosses represent the average (at steady-state) NMDS value of each glacial deposit type, while the length of the crosses represent the standard deviation. (B) NMDS coordination (as in A) with vectors indicating significant DOM characteristics of the medium. The direction of the vector indicate an increase of the variable and the length of the vector indicate the strength of the association.

rates were highest in FLU. For the second period (>12 h) FLU and LAC had comparable lower respiration rates while MOR maintained a respiration rate which was comparable to the initial stage (<12) (ANOVA, $p > 0.05$).

Bacterial carbon production rates were highest for the MOR culture, followed by the LAC and FLU cultures (**Table 2**) and the BCP rates differed significantly between all the glacial deposit types (Kruskal–Wallis, $p < 0.001$). Based on the change in DOC concentration (see section "Alteration of DOM in the Cultures") between medium and culture, the DOC uptake over 24 h (**Table 2**) was found to be similar between LAC and MOR, but significantly lower in FLU (ANOVA, $p < 0.01$, respectively). Bacterial growth efficiency (**Table 2**) was found to be significantly higher in the MOR culture (ANOVA, $p < 0.001$), while the BGE in the FLU and LAC cultures did not differ significant between each other (ANOVA, $p = 0.14$).

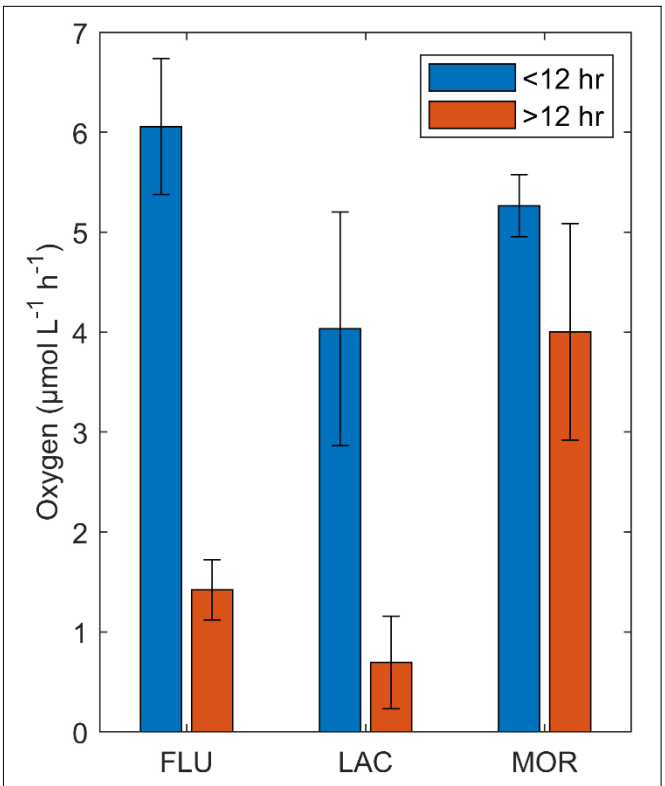


FIGURE 8 | Respiration rates (oxygen µmol L^{-1} h^{-1}) for each glacial deposit type, divided into two different categories <12 h (blue bar) and >12 h (red bar). Error bars represent standard deviation.

DISCUSSION

Characteristics of DOM in Media

Our results show that the composition and biodegradability of DOM differs between post-glacial landscape units and that these differences are related to the glacial processes. In agreement with our results, a recent study also found a coupling between permafrost soil formation and DOM character for different permafrost end-member types (tills, diamicton, lacustrine, peat, and Yedoma deposits) (MacDonald et al., 2021). In addition, different soil-forming factors such as ice content, permafrost extent and parent material (epigenetic vs. syngenetic formation) shape the biogeochemical response to permafrost thaw in aquatic systems (Tank et al., 2020). Factors such as grain size andthe amount of minerals can also influence the OM content (Palmtag and Kuhry, 2018; Opfergelt, 2020). These findingsunderline that soil formation and erosion conditions play an important role when looking at DOM release due to coastal erosion, but at the same time highlight the complexity of studying biodegradation related to permafrost thaw. The differences in soil formation history, OC content, and grain size therefore likely

help explain some of the subtle differences we observed in DOM characteristics between our glacial deposit types.

The DOC:DON ratios (17–25) of the DOM in the media in this experiment indicate a high bioavailability of freshly eroded coastal permafrost soil. Ratios were on the whole lower than those for Arctic rivers [>50, (Dittmar and Kattner, 2003; Holmeset al., 2012)] but comparable to four different Arctic lagoonsand coastal waters (16–19) near our sampling site (Dunton and Crump, 2014) as well as thermokarst feature outflows (19.5) (Abbott et al., 2014). Similarly to these studies, Dittmar and Kattner has reported an average DOC:DON ratio for the ACZ around the Laptev Sea of 21.4 (150 μM DOC and 7 μM DON). $SUVA_{254}$ is correlated to the aromatic content of DOM (Weishaar et al., 2003) and $SUVA_{254}$ from permafrost-derived DOM are expected to vary between 0.6 and 4.5 L mg C^{-1} m^{-1} (Ward and Cory, 2015; Raudina et al., 2017; Wickland et al., 2018; Coch et al., 2019; Fouché et al., 2020) depending on the OM content, with mineral soil types posing lower $SUVA_{254}$ values than organic soil types. In our study, the average $SUVA_{254}$ valueacross the three glacial deposit types was found to be 1.66 Lmg C^{-1} m^{-1} and therefore fits with the reported average of 1.82 L mg C^{-1} m^{-1} obtained from 221 Arctic Canadian soil samples including active layer and permafrost layer (Fouché et al., 2020). Since the contribution of polyphenolic and condensed aromatic compounds leads to lower degradability of organic matter (Textor et al., 2019), we could therefore expect a higher bioavailability in MOR compared to FLU and LAC, based on the $SUVA_{254}$ values. As with the DOC:DON ratios, the $SUVA_{254}$ values are lower than that reported for DOM in Arctic rivers (Walker et al., 2013; Mann et al., 2016; O'Donnell et al., 2016; Coch et al., 2019). DOM is transformed during the transport along the river (Drake et al., 2018; Zhang et al., 2020) and according to the River Continuum Concept it can be expected that labile DOM from the catchment would be rapidly removed, while more recalcitrant DOM would be transported downstream (Vannote et al., 1980). This might explain the higher $SUVA_{254}$ and DOC:DON ratio in Arctic rivers compared to the media used in the experiment.

Slope ratio has shown to be inversely correlated to molecular weight, with low values (<1.0) typical for terrestrial DOM (Helms et al., 2008). DOM extracted from Canadian Arctic soil types expect to have S_r values that range between 0.58 and 3.24 (Fouché et al., 2020). Ward and Cory (2015) demonstrated that the CDOM in the active layer and permafrost layer from soilsin Alaska had S_r values of 0.72 and 0.96, respectively, while Wickland et al. (2018) reported S_r values ranging between 0.73 and 1.13 across the active layer and permafrost layer for three other soil types in Alaska (Orthels, Histels, and Turbels). S_r valuesincrease as DOM is processed in natural settings, particularly via photodegradation (Helms et al., 2008). However, since our media has been kept in dark during the whole experiment, theS_r values in this study relate to the composition of CDOM rather than processing. In addition, it has been shown that lower spectral slopes in the UVA area for Arctic coastal water is not correlated with photodegradation (Juhls et al., 2019), and thatthe lower spectral slope for UVA instead likely reflects lability of DOM (Matsuoka et al., 2015). Therefore, we conclude that the significant differences in S_r values between the three glacial deposit types underline differences in the DOM composition. The higher UVA spectral slope for FLU medium compared to LAC and MOR media (**Table 1**) indicates a lower bioavailability of the DOM pool in this glacial deposit type and fits with results from the BGE values (**Table 2**).

An important difference between the DOM from natural coastal erosion and the DOM in this experiment is the fact that the media was autoclaved. Autoclaving might cause changes in DOM composition due to hydrolysis and denaturation of various compounds and colloids (Dill and Shortle, 1991; Druart and De Wulf, 1993) and, therefore, also changes in bioavailability. However, all the parameters reported above were within natural ranges for soils found in the study region. In addition, it has been shown that even if autoclaving changed DOC in an unpredictable manner, it did not cause a convergence of the DOC pool from different lakes (Andersson et al., 2018). This means that although DOC is not identical to the initial conditions after autoclaving, the diversity of DOC is preserved.

Optical Signature of Microbial Degradation

The microbial degradation of the DOM in the media imparted an optical signature on the absorption and fluorescence properties of the DOM. Despite the fact that the initial spectral characteristics of the DOM in the media were only subtly different between glacial deposit types, the results show significant differences in DOM turnover. The high utilization rate of permafrost-derived DOM has been correlated to the relative high abundance of hydrogen-rich compounds, such as aliphatic molecules (amino acids, peptides, and protein) and carbohydrates (Spencer et al., 2015; Stapel et al., 2017; Textor et al., 2018, 2019). Lower degradability of organic matter in some soils has on the other hand been correlated with a greater contribution of polyphenolic and condensed aromatic compounds, often linked to decomposition processes of the overlying vegetation during unfrozen periods (Textor et al., 2019). Although harder for bacteria to utilize, compounds such as lignin phenols and related poly-phenolic compounds, can be metabolized or transformed into other compounds (Fasching et al., 2014).

Even though the highest amount of DOC was degraded in the MOR culture (**Figure 2**), the optical signature indicated a net production of CDOM and FDOM (**Figures 4**, **5**). Production of fluorescence peaks B and T in coastal waters are known to be correlated to amino acids produced by bacterial communities (Yamashita and Tanoue, 2003). This production of CDOM and FDOM, together with the high DOC removal, suggests that the DOM derived from the MOR deposit type contains a high amount of colorless labile DOM compounds (less conjugated aliphatic molecules). Since these compounds are easier for bacteria to utilize (Berggren et al., 2010), CDOM and FDOM would not be degraded as long as these labile DOM compounds are available. The high amount of colorless DOM can probably be explained by the glacial formation process behind MOR. Sincethe MOR deposit type was not submerged in water during glacialprocesses (Krzic et al., 2010), little prior decomposition of these

less conjugated aliphatic compounds has occurred. The upper permafrost layer of MOR may have thawed during the Holocene Thermal Maximum, where active layer depths reached more than 1 m (Burn, 1997), which could have resulted in intensive carbon degradation under aerobic conditions as a topographic gradient remained. However, this assumption remains speculative for our sampled sites.

From the optical signature (**Figures 4**, **5**), the FLU culture showed degradation of CDOM and FDOM compounds across the whole absorbance spectrum and fluorescence EEM. This suggests that only a very small amount of colorless bioavailable DOM was present and that the microbial communitydegraded the bioavailable CDOM and FDOM, such as labile conjugated aliphatic and aromatic compounds, immediately. This assumption fits well with the FLU deposit type, which potentially lacks less conjugated and colorless labile aliphatic compounds, previously leached from vegetation into the active layer (Textor et al., 2019). The lack of these compounds could be due to degradation when the FLU deposit type was submerged under stagnant waters during glacial processes in the Holocene Thermal Maximum (Krzic et al., 2010; Schirrmeister et al., 2011), thereby leaving a higher abundance of more conjugated aliphatic and aromatic compounds behind when the soil permanently refroze after this period.

For the LAC culture, we observed a production of peak B (**Figure 5**), which could indicate that the DOM derived from the LAC deposit type was transformed into FDOM compounds upon microbial degradation (Yamashita and Tanoue, 2003; Fasching et al., 2014). However, in contrast to the MOR culture, theLAC culture also degrades CDOM and FDOM compounds (**Figures 4**, **5**), such as labile conjugated aliphatic and aromatic compounds, probably as the competition for the colorless labile DOM increases. Similar to the FLU deposit type, theLAC deposit type was also submerged in water during glacial processes (Krzic et al., 2010), resulting in the degradation ofless conjugated and colorless aliphatic compounds prior to our sampling (Walter Anthony et al., 2018). In contrast to the fluvial environment, the lacustrine environment has probably allowed a production of colorless aliphatic compounds (Meyers and Ishiwatari, 1995; Schirrmeister et al., 2011), such as OM excreted from phytoplankton and heterotrophic species which can be very labile compounds and often support bacterial growth(Rosenstock and Simon, 2001; Kinsey et al., 2018).

Our results suggest that the DOM composition in Cryosols with a different glacial geomorphic history, induce marine microbial communities to impart different optical degradation signatures, ultimately indicating differences in biodegradability among the glacial deposit types. The results also indicate that DOM in the media is more bioavailable than riverine terrestrial DOM, since several studies found no to very low degradation by marine microbial communities as the riverine terrestrial DOM enters coastal waters (Köhler et al., 2003; Amon andMeon, 2004; Herlemann et al., 2014; Blanchet et al., 2017). Large rivers have already lost most of the labile ancient soil DOM components at the time the material reach the ACZ, as degradation occurs mostly in uplands and headwaters (Drakeet al., 2015; Spencer et al., 2015). Nonetheless, it can therefore be argued that DOM released from coastal erosion of Cryosols will have a larger impact on the coastal environment in the ACZ than that of riverine DOM.

Substrate Driven Community Differences

The observed distinct patterns in BCC between the three glacial deposit types, used in this experiment, reflect the subtle underlying differences in DOM quality (**Figure 7** and **Supplementary Table 1**). The grouping of the BCC across replicates (**Figure 7**) within each soil treatment indicates a systematic DOM control on BCC development from a common marine inoculum community. Interestingly, the community composition in the MOR culture was not positively correlated to any of the DOM characteristics included in the analysis which might indicate that this community is selected by the colorless DOC dominating the MOR medium. This agrees with the negative correlation to peak T and $SUVA_{254}$ for MOR soil, which indicates a response to less degraded DOM pool with a lower aromaticity. In contrast, the BCC of FLU and LAC cultures were positively correlated to one or several of the DOM characteristics (**Table 1**). The positive correlation to peak B in the FLU medium indicate a response to a higher relative amount of protein-like compounds, whereas the positive correlation to peak A, peak C, peak M, and $SUVA_{254}$ in the LAC soil show a response to humic-like compounds (Coble, 2007; Stedmon and Nelson, 2015).

Common for all of the three different cultures was the reduction in species richness between start and steady state,which essentially indicates that cultivation in the chemostat selected bacterial communities that were best suited to the specific DOM derived from the glacial deposit types. The rise anddominance of *Gammaproteobacteria* and *Alphaproteobacteria* in these chemostat cultures agrees with results from regrowth batch experiments (Sipler et al., 2017; Müller et al., 2018). However, *Gammaproteobacteria* are commonly found to dominate biodegradation experiments, as these bacterial taxaare known to be opportunistic with high growth rate and ability to exploit available DOM (Herlemann et al., 2014). This was clearly the case for the MOR culture (**Figure 6**) where the highest DOC uptake and respiration rates were measured (**Figure 8**), coupled with CDOM and FDOM production (**Figures 4**, **5**). Combined these results indicate that moraine soil DOM containscolorless labile DOM which is rapidly degraded and supporting the development of the *Gammaproteobacteria* community. Inthe FLU and LAC cultures, we instead observed a dominanceof *Alphaproteobacteria* (**Figure 6**). This difference in BCC was paralleled with the notable difference in CDOM and FDOM signatures imparted by the communities (Mantel tests, **Supplementary Table 3**). In FLU and LAC cultures, there were a significant removal of CDOM and FDOM (**Figures 4**, **5**). Although both dominated by *Alphaproteobacteria*, the FLU and LAC communities were also distinct (**Figure 7**) and this apparently influenced the extent of CDOM and FDOM removal (**Figures 4**, **5**).

Alphaproteobacterial reads were dominated by the order *Rhodobacterales* in cultures from all glacial deposit types, whereas *Oceanospirillales* was the dominant order among Gammaproteobacterial reads. However, the difference in BCC

was verified at the genus level where several ASVs differed significantly in abundance between the three glacial deposit types (tagwise dispersion analysis). There was significant higher abundance of ASVs belonging to the genus *Sulfitobacter* in the FLU culture, and members from this genus have been isolated from similar environments (Park et al., 2019). The first isolate and type specie, *Sulfitobacter pontiacus*, is a sulfur-oxidizing chemoheterotrophic bacteria which utilizes mainly carboxyl and amino acids (Sorokin, 1995). In the LAC cultures several of the ASVs with significantly higher abundance belonged to the genus *Polaribacter* and *Pacificibacter*. *Polaribacter* belongs to *Flavobacteria* (*Bacteroidetes*) which have been ascribed to act as degraders of high molecular weight OM, such as proteins and carbohydrates (Thomas et al., 2011). The type specie for *Pacificibacter*, *Pacificibacter maritimus*, was isolated from shallow marine sediments and utilizes mostly sugars, amino acids and a few carboxylic acids (Romanenko et al., 2011). And last, in the MOR culture several of the ASVs belonged to the genera *Amphritea* and *Marinomonas*. The majority of the members of the genus *Amphritea* are closely associated with living marine organisms, however, some members of the genus has been isolated from marine sediments (Miyazaki et al., 2008). Species within *Amphritea* oxidize various sugars and carboxylic acids (Gärtner et al., 2008). Members of the genus *Marinomonas* have a widespread distribution in marine environments and have for example been found in seawater (Yoon et al., 2005), sea ice (Zhang et al., 2008) and seafloor sediment (Romanenko et al., 2011). Based on the characterization of *Marinomonas polaris* and *Marinomonas arctica* the species utilize sugars, amino acids, and sugar alcohols, but not complex carboxylic acids and aromatic compounds (Gupta et al., 2006; Zhang et al., 2008).

Although it is hard to link these ASVs to specific characteristics of the DOM, these analyses show that the presence and abundance of specific phylotypes might determine the utilization of DOM. Our results are therefore in line withthe growing number of studies linking community composition to OM lability (Cottrell and David, 2003; Gómez-Consarnauet al., 2012; Nelson and Carlson, 2012; Logue et al., 2016;Balmonte et al., 2019) and shows that the bioavailability should be seen as an interaction between the chemical composition of DOM and the metabolic capacity of the microbial community (Nelson and Wear, 2014).

Microbial Carbon Processing

Climate change will intensify erosion of Arctic coasts (Günther et al., 2015; Hoque and Pollard, 2016; Obu et al., 2016; Couture et al., 2018; Jones et al., 2018) but the ultimate fate of this carbon source is still unknown (Fritz et al., 2017). Understanding how DOC from different glacial deposit types will be mobilized by bacteria in the ACZ could eventually help us quantify the fate of carbon export from eroding soils along Arctic coastlines. We acknowledge the important difference between natural coastal erosion and our experiment, where coastal erosion will add both OM and *in situ* microbial communities to the marine water. Recent studies along aquatic continua have shown that microbial communities in lakes and streams contain organisms with terrestrial origins (Crump et al., 2012; Ruiz-González et al., 2015; Hauptmann et al., 2016). However, it is not clear whether bacteria transported with soil will be active and thrive in sea water since salinity is one of the strongest environmental filters (Langenheder et al., 2003). In addition to 'species sorting' which probably will take place in response to this chemical limit (Van der Gucht et al., 2007) there will also be priority effectsfor well established communities (Svoboda et al., 2018). We therefore believe that most of the carbon reaching the ACZdue to coastal erosion will be processed by the marine bacterial community.

In this study we demonstrate that not only are colorless DOM compounds being mobilized by marine microbial communities when Cryosols are released into the ACZ, but also that CDOM and FDOM compounds are bioavailable to marine microbial communities. The DOC removal by marine microbial degradation is the sum of carbon used for bacterial production and carbon released through respiration (**Table 2**). The bacterial production achieved by the chemostat cultures is dependent on the dilution rate, the quality of substrate (DOM bioavailability) and the ability of the community to utilize the substrate (Del Giorgio and Cole, 1998). As the dilution rates were set constant, the differences observed here reflect a combinationof community composition and DOM character. The FLU culture had the lowest removal of DOC (**Figure 2** and **Table 2**) and lowest BCP. MOR and LAC cultures achieved approximately 3–4 times higher BA than FLU and abundances were higher than those often achieved in bottle experiment (Sipler et al., 2017; Müller et al., 2018), where supply of substrateis limited. Although the DOC removal was similar betweenthe MOR and LAC cultures, BCP was two times higher inthe MOR cultures.

Differences in BCP will have consequences for bacterial growth efficiency (BGE) which is a measurement of how efficiently carbon is converted into biomass (Del Giorgio and Cole, 1998). BGE has been shown to be strongly correlated with the composition of terrestrial DOM (Berggren et al., 2007). BGE in our study varied between 13 and 66% with highest values in the MOR culture and lowest in the FLU culture. Estimates of BGE for natural aquatic bacterial communities range from 5 to 60% with the highest values usually foundin coastal waters and estuaries (Del Giorgio and Cole, 1998). However, in the Arctic region other studies have found the BGE to be 19.5% within Arctic Rivers and between 7 and 10% within Arctic Fjords (Middelboe et al., 2012; Paulsen et al.,2017), while it has been found to be lower in Arctic Oceanwith 6.3% in the Fram Strait (Kritzberg et al., 2010) and 2.2%in the Chucki Sea (Cota et al., 1996). The fact that the BGE results from our study exceeds what has been previously found within other aquatic ecosystems throughout the Arctic Ocean, therefore suggest that the DOC from freshly exposed Cryosols could be one of the most labile sources of carbon in this region. In addition, the high variation in BGE between glacial deposit types suggest that the lability of DOC depends heavilyon the specific deposit type being eroded. The increased rateof coastal erosion and thereby release of highly bioavailableDOM can therefore have a huge impact on the marine microbial communities in the ACZ.

Several studies have shown that a potentially large amount of CO_2 is released during coastal erosion due to microbial processes (Vonk et al., 2012; Semiletov et al., 2013; Tanski et al., 2019). Our results on BCP and respiration show that carbon processing by the marine microbial communities will lead to different fates of the DOC depending on the glacial deposit type. With a lower BGE, lower amounts of bioavailable DOC are stored as biomass in the marine microbial communities and more of the carbon may be respired to the atmosphere as CO_2, as seen for the FLU deposit type. However, the majority of DOC was indeed refractory to biodegradation. This refractory DOC may be exported further out into ACZ where it may end up being buried in sediments or stored in the deep ocean.

Summary of Findings and Future Implications

Coastal erosion in the Arctic is intensifying and DOM from Cryosols will become an even more important source in the ACZ in the future. Here we show that the composition and biodegradability of DOM differs between post-glacial landscape units and that the differences in DOM composition and biodegradability are related to the glacial processes.

The three different DOM sources (FLU, LAC and MOR) supported the development of three different marine microbial communities, which was especially clear when comparing the MOR deposit type to FLU and LAC. These findings indicate a clear substrate-driven control on marine microbial community composition, especially where the input of organic carbon and DOM in the ACZ is dominated by releasefrom coastal erosion of Cryosols. The bacterial communitiesimparted different spectral fingerprints on the absorption spectra and fluorescence EEMs of the DOM. Based onthese fingerprints, we show that both colorless DOM and labile CDOM and FDOM fractions are being degraded. Also, the spectral results suggest that the more refractory CDOM and FDOM pool can be associated with CDOM and FDOM found in the open ocean after passing the ACZ.In addition, we show that absorbance at 330 nm could to be a proxy of microbial degradation of CDOM, especially produced in FLU and LAC deposit types during the Holocene Thermal Maximum.

The chemostat approach applied here provides a simulation of the constant substrate supply to coastal waters that can occur during summer open water conditions with maximum permafrost erosion rates. However, the results from thisexperiment will not fully represent what is happening in the natural environment. Important differences include the higher-
than-natural incubation temperature at 20 ˚C, which is standard for bioavailability studies, but might affect both the activityand composition of the bacterial community (Pomeroy and Wiebe, 2001; Adams et al., 2010; Sjöstedt et al., 2012a), togetherwith autoclaving of the media, which can cause changes inthe bioavailability and composition of the DOM. However, chemostats has been suggested to be the method that most closely resembles the growth conditions bacteria encounter in natural systems (Kovárová-Kovar and Egli, 1998) and the reproducibility between the replicate chemostats provide confidence in the robustness of this method. It is therefore clear that DOM quality influences the BCC, which in turn also affects the net effect of DOM degradation. Moraine deposit type will result in the net production of CDOM and FDOM in coastal waters, while the deposits types that are formed in aquatic environments, such as FLU and LAC, will lose CDOM andFDOM as it passes through coastal waters. These findings suggest a continuum, where the presence of ancient colorless labile DOM supports a rapidly growing community and a net production of CDOM and FDOM, which in turn can be degraded by other members of the bacterial communities with the capacity to degrade CDOM and FDOM.

To achieve a better understanding of carbon turnover from coastal erosion of Cryosols in the ACZ and its effects on climate and ecology in the future, more studies including qualitative measurements on DOM, such as absorbance and fluorescence spectroscopy, coupled with both bacterial and phytoplankton community assessment should be performed. Nevertheless, it is important to note that a large proportion of bioavailable DOM in these systems may not be captured and characterized using optical measurements.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found online using the following link: https://doi.org/10.11583/DTU.14113250.v1. The names of accession number(s) can be found below: https: //www.ncbi.nlm.nih.gov/, PRJNA675030.

AUTHOR CONTRIBUTIONS

AB, CS, and JS planned and performed the experiments. AM, NS, GT, and JV led sampling site selection and coordinatedand performed field sampling. AB, CS, JS, and JC analyzed samples and performed data analysis. AB, CS, and JS wrotethe manuscript. All authors commented on the manuscript and contributed to the interpretation and discussion of the results.

FUNDING

This publication is part of the Nunataryuk project. Theproject has received funding under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 773421. Part of this research was supported by Japan Aerospace Exploration Agency (JAXA) Global Change Observation Mission-Climate (GCOM-C) to AM (contract #20RT000350), Independent Research Fund Denmark (9040-00266B) awarded to CS, and by the Swedish Research Council (VR, grant 2015-00188) to JS, and the Natural Sciences and Engineering Research Council of Canada (Discovery program, RGPIN-2020-06874) to JC.

ACKNOWLEDGMENTS

We acknowledge colleagues in the EU Horizon 2020 Nunataryuk project (grant no. 773421) who contributed to the sampling and logistics and colleagues at Plateforme d'Analyses Génomiques (IBIS, Université Laval, Quebec City, Canada) for performing the Illumina sequencing. H. Lantuit, M. Fritz, and G. Hugelius are thanked for providing logistical and technical support. We also thank Lea Tolstrup for assistance during the chemostat experiment. Finally we thank the editor and the three reviewers for their valuable comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/feart.2021.640580/full#supplementary-material

REFERENCES

Abbott, B. W., Larouche, J. R., Jones, J. B., Bowden, W. B., and Balser, A. W. (2014). Elevated dissolved organic carbon biodegradability from thawing and collapsing permafrost. *J. Geophys. Res. Biogeosci.* 119, 2049–2063. doi: 10.1002/2014JG002678

Adams, H. E., Crump, B. C., and Kling, G. W. (2010). Temperature controls on aquatic bacterial production and community dynamics in arctic lakes and streams. *Environ. Microbiol.* 12, 1319–1333. doi: 10.1111/j.1462-2920.2010.02176.x

Amon, R. M. W., and Meon, B. (2004). The biogeochemistry of dissolved organic matter and nutrients in two large Arctic estuaries and potential implications for our understanding of the Arctic Ocean system. *Mar. Chem.* 92, 311–330. doi: 10.1016/j.marchem.2004.06.034

Andersson, M. G. I., Catalán, N., Rahman, Z., Tranvik, L. J., and Lindström, E. S. (2018). Effects of sterilization on dissolved organic carbon (DOC) composition and bacterial utilization of DOC from lakes. *Aquat. Microb. Ecol.* 82, 199–208. doi: 10.3354/ame01890

Arrigo, K. R., and Brown, C. W. (1996). Impact of chromophoric dissolved organic matter on UV inhibition of primary productivity in the sea. *Mar. Ecol. Prog. Ser.* 140, 207–216. doi: 10.3354/meps140207

Balmonte, J. P., Buckley, A., Hoarfrost, A., Ghobrial, S., Ziervogel, K., Teske, A., et al. (2019). Community structural differences shape microbial responses to high molecular weight organic matter. *Environ. Microbiol.* 21, 557–571. doi: 10.1111/1462-2920.14485

Belicka, L. L., Macdonald, R. W., and Harvey, H. R. (2002). Sources and transport of organic carbon to shelf, slope, and basin surface sediments of the Arctic Ocean. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 49, 1463–1483. doi: 10.1016/S0967-0637(02)00031-6

Berggren, M., Laudon, H., Haei, M., Ström, L., and Jansson, M. (2010). Efficient aquatic bacterial metabolism of dissolved low-molecular-weight compounds from terrestrial sources. *ISME J.* 4, 408–416. doi: 10.1038/ismej.2009.120

Berggren, M., Laudon, H., and Jansson, M. (2007). Landscape regulation of bacterial growth efficiency in boreal freshwaters. *Glob. Biogeochem. Cycles* 21:GB4002. doi: 10.1029/2006GB002844

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., et al. (2019). Permafrost is warming at a global scale. *Nat. Commun.* 10:264. doi: 10.1038/s41467-018-08240-4

Blanchet, M., Pringault, O., Panagiotopoulos, C., Lefèvre, D., Charrière, B., Ghiglione, J.-F., et al. (2017). When riverine dissolved organic matter (DOM) meets labile DOM in coastal waters: changes in bacterial community activity and composition. *Aquat. Sci.* 79, 27–43. doi: 10.1007/s00027-016-0477-0

Burn, C. R. (1997). Cryostratigraphy, paleogeography, and climate change during the early Holocene warm interval, western Arctic coast, Canada. *Can. J. Earth Sci.* 34, 912–925. doi: 10.1139/e17-076

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Canelhas, M. R., Eiler, A., and Bertilsson, S. (2016). Are freshwater bacterioplankton indifferent to variable types of amino acid substrates? *FEMS Microbiol. Ecol.* 92:fiw005. doi: 10.1093/femsec/fiw005

Cauwet, G. (1999). ""Determination of dissolved organic carbon and nitrogen by high temperature combustion," in *Methods of Seawater Analysis*, eds K. Grasshoff, K. Kremling, and M. Ehrhardt (Weinheim: Wiley-VCH Verlag GmbH), 407–420. doi: 10.1002/9783527613984.ch15

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

Coble, P. G. (2007). Marine optical biogeochemistry: the chemistry of ocean color. *Chem. Rev.* 107, 402–418. doi: 10.1021/cr050350

Coch, C., Juhls, B., Lamoureux, S. F., Lafreniére, M. J., Fritz, M., Heim, B., et al. (2019). Comparisons of dissolved organic matter and its optical characteristics in small low and high Arctic catchments. *Biogeosciences* 16, 4535–4553. doi: 10.5194/bg-16-4535-2019

Colatriano, D., Tran, P. Q., Guéguen, C., Williams, W. J., Lovejoy, C., and Walsh, D. A. (2018). Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria. *Commun. Biol.* 1:90. doi: 10.1038/s42003-018-0086-7

Cota, G. F., Pomeroy, L. R., Harrison, W. G., Jones, E. P., Peters, F., Sheldon, W. M., et al. (1996). Nutrients, primary production and microbial heterotrophy in the southeastern Chukchi Sea:Arctic summer nutrient depletion and heterotrophy. *Mar. Ecol. Prog. Ser.* 135, 247–258. doi: 10.3354/meps1 35247

Cottrell, M. T., and David, K. L. (2003). Contribution of major bacterial groups to bacterial biomass production (thymidine and leucine incorporation) in the Delaware estuary. *Limnol. Oceanogr.* 48, 168–178. doi: 10.4319/lo.2003.48.1.0168

Couture, N. J., Irrgang, A., Pollard, W., Lantuit, H., and Fritz, M. (2018). Coastal erosion of permafrost soils along the yukon coastal plain and fluxes of organic carbon to the canadian beaufort sea. *J. Geophys. Res. Biogeosci.* 123, 406–422. doi: 10.1002/2017JG004166

Crump, B. C., Amaral-Zettler, L. A., and Kling, G. W. (2012). Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils. *ISME J.* 6, 1629–1639. doi: 10.1038/ismej.2 012.9

Del Giorgio, P. A., and Cole, J. J. (1998). Bacterial growth efficiency in natural aquatic systems. *Annu. Rev. Ecol. Syst.* 29, 503–541. doi: 10.1146/annurev.ecolsys.29.1.503

Dill, K. A., and Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* 60, 795–825. doi: 10.1146/annurev.bi.60.070191.004051

Dittmar, T., and Kattner, G. (2003). The biogeochemistry of the river and shelf ecosystem of the Arctic ocean: a review. *Mar. Chem.* 83, 103–120. doi: 10.1016/S0304-4203(03)00105-1

Dou, F., Ping, C.-L., Guo, L., and Jorgenson, T. (2008). Estimating the impact of seawater on the production of soil water-extractable organic carbon during coastal erosion. *J. Environ. Qual.* 37, 2368–2374. doi: 10.2134/jeq2007.0403

Drake, T. W., Guillemette, F., Hemingway, J. D., Chanton, J. P., Podgorski, D. C., Zimov, N. S., et al. (2018). The ephemeral signature of permafrost carbon in an arctic fluvial network. *J. Geophys. Res. Biogeosci.* 123, 1475–1485. doi: 10.1029/2017JG004311

Drake, T. W., Wickland, K. P., Spencer, R. G. M., McKnight, D. M., and Striegl, R. G. (2015). Ancient low-molecular-weight organic acids in permafrost fuel rapid carbon dioxide production upon thaw. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13946–13951. doi: 10.1073/pnas.1511705112

Druart, P., and De Wulf, O. (1993). Activated charcoal catalyses sucrose hydrolysis during autoclaving. *Plant Cell Tiss. Organ Cult.* 32, 97–99. doi: 10.1007/BF00040122

Dunton, K., and Crump, B. (2014). *Collaborative Research: Terrestrial Linkages to Microbial and Metazoan Communities in Coastal Ecosystems of the Beaufort Sea*. Santa Barbara, CA: Arctic Data Center.

Environment Canada (2016). *Canadian Climate Normals*. Available online at: https://climate.weather.gc.ca/climate_normals/ (accessed February 1, 2021).

Fasching, C., Behounek, B., Singer, G. A., and Battin, T. J. (2014). Microbial degradation of terrigenous dissolved organic matter and potential consequences for carbon cycling in brown-water streams. *Sci. Rep.* 4:4981. doi: 10.1038/srep04981

Fouché, J., Christiansen, C. T., Lafrenière, M. J., Grogan, P., and Lamoureux, S. F. (2020). Canadian permafrost stores large pools of ammonium and optically distinct dissolved organic matter. *Nat. Commun.* 11:4500. doi: 10.1038/s41467-020-18331-w

Fritz, M., Opel, T., Tanski, G., Herzschuh, U., Meyer, H., Eulenburg, A., et al. (2015). Dissolved organic carbon (DOC) in Arctic ground ice. *Cryosphere* 9, 737–752. doi: 10.5194/tc-9-737-2015

Fritz, M., Vonk, J. E., and Lantuit, H. (2017). Collapsing arctic coastlines. *Nat. Clim. Chang.* 7, 6–7. doi: 10.1038/nclimate3188

Fritz, M., Wetterich, S., Schirrmeister, L., Meyer, H., Lantuit, H., Preusser, F., et al. (2012). Eastern Beringia and beyond: late Wisconsinan and holocene landscape dynamics along the Yukon Coastal Plain, Canada. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 319–320, 28–45. doi: 10.1016/j.palaeo.2011.12.015

Gärtner, A., Wiese, J., and Imhoff, J. F. (2008). *Amphritea atlantica* gen. nov., sp. nov., a gammaproteobacterium from the Logatchev hydrothermal vent field. *Int. J. Syst. Evol. Microbiol.* 58, 34–39. doi: 10.1099/ijs.0.65234-0

Gasol, J. M., and Del Giorgio, P. A. (2000). Using flow cytometry for counting natural planktonic bacteria and understanding the structure of planktonic bacterial communities. *Sci. Mar.* 64, 197–224. doi: 10.3989/scimar.2000.64n2197

Gómez-Consarnau, L., Lindh, M. V., Gasol, J. M., and Pinhassi, J. (2012). Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environ. Microbiol.* 14, 2361–2378. doi: 10.1111/j.1462-2920.2012.02804.x

Günther, F., Overduin, P. P., Yakshina, I. A., Opel, T., Baranskaya, A. V., and Grigoriev, M. N. (2015). Observing Muostakh disappear: permafrost thaw subsidence and erosion of a ground-ice-rich island in response to arctic summer warming and sea ice reduction. *Cryosphere* 9, 151–178. doi: 10.5194/tc-9-151-2015

Guo, L., and Macdonald, R. W. (2006). Source and transport of terrigenousorganic matter in the upper Yukon River: evidence from isotope (δ^{13} C, Δ^{14}C, and δ^{15} N) composition of dissolved, colloidal, and particulate phases. *Glob. Biogeochem. Cycles* 20:GB2011. doi: 10.1029/2005GB002593

Gupta, P., Chaturvedi, P., Pradhan, S., Delille, D., and Shivaji, S. (2006). *Marinomonas polaris* sp. nov., a psychrohalotolerant strain isolated from coastal sea water off the subantarctic Kerguelen islands. *Int. J. Syst. Evol. Microbiol.* 56, 361–364. doi: 10.1099/ijs.0.63921-0

Hagstrom, Å, Ammerman, J. W., Heinrichs, S., and Azam, F. (1984). Bacterioplankton growth in seawater: II. Growth kinetics and cellular characteristicsin seawater cultures. *Mar. Ecol. Prog. Ser.* 18, 41–48.

Hansen, H. P., and Koroleff, F. (1999). "Determination of nutrients," in *Methods of Seawater Analysis*, eds K. Grasshoff, K. Kremling, and M. Ehrhardt (Weinheim: Wiley-VCH Verlag GmbH), 159–228. doi: 10.1002/9783527613984.ch10

Hauptmann, A. L., Markussen, T. N., Stibal, M., Olsen, N. S., Elberling, B., Baelum, J., et al. (2016). Upstream Freshwater and Terrestrial Sources Are Differentially Reflected in the Bacterial Community Structure along a Small Arctic River and its estuary. *Front. Microbiol.* 7:1474. doi: 10.3389/fmicb.2016.01474

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

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41

Herlemann, D. P. R., Manecki, M., Meeske, C., Pollehne, F., Labrenz, M., Schulz-Bull, D., et al. (2014). Uncoupling of bacterial and terrigenous dissolved organic matter dynamics in decomposition experiments. *PLoS One* 9:e93945. doi: 10.1371/journal.pone.0093945

Heslop, J. K., Winkel, M., Walter Anthony, K. M., Spencer, R. G. M., Podgorski, D. C., Zito, P., et al. (2019). Increasing organic carbon biolability with depth in yedoma permafrost: ramifications for future climate change. *J. Geophys. Res. Biogeosci.* 124, 2021–2038. doi: 10.1029/2018JG004712

Holmes, R. M., McClelland, J. W., Peterson, B. J., Tank, S. E., Bulygina, E., Eglinton, T. I., et al. (2012). Seasonal and Annual Fluxes of Nutrients and Organic Matter from large rivers to the Arctic Ocean and surrounding Seas. *Estuar. Coasts* 35, 369–382. doi: 10.1007/s12237-011-9386-6

Hoque, M. A., and Pollard, W. H. (2016). Stability of permafrost dominated coastal cliffs in the Arctic. *Polar Sci.* 10, 79–88. doi: 10.1016/j.polar.2015.10.004

Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C. L., et al. (2014). Estimated stocks of circumpolar permafrost carbon with quantified' ' uncertainty ranges and identified data gaps. *Biogeosciences* 11, 6573–6593. doi: 10.5194/bg-11-6573-2014

Jobbágy, E. G., and Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* 10, 423–436.

Jones, B. M., Farquharson, L. M., Baughman, C. A., Buzard, R. M., Arp, C. D., Grosse, G., et al. (2018). A decade of remotely sensed observations highlight complex processes linked to coastal permafrost bluff erosion in the Arctic. *Environ. Rese. Lett.* 13:115001. doi: 10.1088/1748-9326/aae471

Juhls, B., Overduin, P. P., Hölemann, J., Hieronymi, M., Matsuoka, A., Heim, B., et al. (2019). Dissolved organic matter at the fluvial–marine transition in the Laptev Sea using in situ data and ocean colour remote sensing. *Biogeosciences* 16, 2693–2713. doi: 10.5194/bg-16-2693-2019

Kaiser, K., and Guggenberger, G. (2000). The role of DOM sorption to mineral surfaces in the preservation of organic matter in soils. *Org. Geochem.* 31, 711–725. doi: 10.1016/S0146-6380(00)00046-2

Kawahigashi, M., Kaiser, K., Rodionov, A., and Guggenberger, G. (2006). Sorption of dissolved organic matter by mineral soils of the Siberian forest tundra. *Glob. Chang. Biol.* 12, 1868–1877. doi: 10.1111/j.1365-2486.2006.01203.x

Kinsey, J. D., Corradino, G., Ziervogel, K., Schnetzer, A., and Osburn, C. L. (2018). Formation of chromophoric dissolved organic matter by bacterial degradation of phytoplankton-derived aggregates. *Front. Mar. Sci.* 4:430. doi: 10.3389/fmars.2017.00430

Köhler, H., Meon, B., Gordeev, V. V., Spitzy, A., and Amon, R. M. W. (2003). "Dissolved organic matter (DOM) in the estuaries of Ob and Yenisei and the adjacent Kara Sea, Russia," in *Proceedings of the Marine Science*, Vol. 6, Amsterdam.

Kovárová-Kovar, K., and Egli, T. (1998). Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. *Microbiol. Mol. Biol. Rev.* 62, 646–666.

Kritzberg, E. S., Duarte, C. M., and Wassmann, P. (2010). Changes in Arctic marine bacterial carbon metabolism in response to increasing temperature. *Polar Biol.* 33, 1673–1682. doi: 10.1007/s00300-010-0799-7

Krzic, M., Watson, K., Grand, S., Bomke, A., Smith, S., Dyanatkar, S., et al. (2010). *Soil Formation and Parent Material. The University of British Columbia, Vancouver, Thompson Rivers University, Kamloops, and Agriculture and Agri-Food Canada, Summerland*. Available online at: https://landscape.soilweb.ca/about/ (accessed November 20, 2020).

Langenheder, S., Kisand, V., Wikner, J., and Tranvik, L. J. (2003). Salinity as a structuring factor for the composition and performance of bacterioplankton degrading riverine DOC. *FEMS Microbiol. Ecol.* 45, 189–202. doi: 10.1016/S0168-6496(03)00149-1

Lawaetz, A. J., and Stedmon, C. A. (2009). Fluorescence intensity calibration using the Raman scatter peak of water. *Appl. Spectrosc.* 63, 936–940. doi: 10.1366/000370209788964548

Lee, S., and Fuhrman, J. A. (1987). Relationships between biovolume and biomass of naturally derived marine bacterioplankton. *Appl. Environ. Microbiol.* 53, 1298–1303. doi: 10.1128/AEM.53.6.1298-1303.1987

Logue, J. B., Stedmon, C. A., Kellerman, A. M., Nielsen, N. J., Andersson, A. F., Laudon, H., et al. (2016). Experimental insights into the importance of aquatic bacterial community composition to the degradation of dissolved organic matter. *ISME J.* 10, 533–545. doi: 10.1038/ismej.20 15.131

MacDonald, E. N., Tank, S. E., Kokelj, S. V., Froese, D. G., and Hutchins, R. H. S. (2021). Permafrost-derived dissolved organic matter composition varies across permafrost end-members in the western Canadian Arctic. *Environ. Res. Lett.* 16:024036. doi: 10.1088/1748-9326/abd971

Mann, P. J., Eglinton, T. I., McIntyre, C. P., Zimov, N., Davydova, A., Vonk, J. E., et al. (2015). Utilization of ancient permafrost carbon in headwaters of Arctic fluvial networks. *Nat. Commun.* 6:7856. doi: 10.1038/ncomms8856

Mann, P. J., Spencer, R. G. M., Hernes, P. J., Six, J., Aiken, G. R., Tank, S. E., et al. (2016). Pan-Arctic trends in terrestrial dissolved organic matter from optical measurements. *Front. Earth Sci.* 4:25. doi: 10.3389/feart.2016.00025

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10. doi: 10.14806/ej.17.1.200

Matsuoka, A., Ortega-Retuerta, E., Bricaud, A., Arrigo, K. R., and Babin, M. (2015). Characteristics of colored dissolved organic matter (CDOM) in the Western Arctic Ocean: Relationships with microbial activities. *Deep Sea Res. 2 Top. Stud. Oceanogr.* 118, 44–52. doi: 10.1016/j.dsr2.2015.02.012

McGuire, A. D., Lawrence, D. M., Koven, C., Clein, J. S., Burke, E., Chen, G., et al. (2018). Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change. *Proc. Natl. Acad. Sci. U.S.A.* 115, 3882–3887. doi: 10.1073/pnas.1719903115

Meyers, P. A., and Ishiwatari, R. (1995). "Organic matter accumulation records in lake sediments," in *Physics and Chemistry of Lakes*, eds A. Lerman, D. M. Imboden, and J. R. Gat (Berlin: Springer), 279–328. doi: 10.1007/978-3-642-85132-2_10

Middelboe, M., Glud, R. N., and Sejr, M. K. (2012). Bacterial carbon cycling in a subarctic fjord: a seasonal study on microbial activity, growth efficiency, and virus-induced mortality in Kobbefjord. *Greenland. Limnol. Oceanogr.* 57, 1732–1742. doi: 10.4319/lo.2012.57.6.1732

Miyazaki, M., Nogi, Y., Fujiwara, Y., Kawato, M., Nagahama, T., Kubokawa, K., et al. (2008). *Amphritea japonica* sp. nov. and *Amphritea balenae* sp. nov., isolated from the sediment adjacent to sperm whale carcasses off Kagoshima. *Jpn. Int. J. Syst. Evol. Microbiol.* 58, 2815–2820. doi: 10.1099/ijs.0.65826-0

Moriarty, D. J. W. (1986). "Measurement of bacterial growth rates in aquatic systems from rates of nucleic acid synthesis," in *Advances in Microbial Ecology Advances in Microbial Ecology*, ed. K. C. Marshall (Boston, MA: Springer), 245–292. doi: 10.1007/978-1-4757-0611-6_6

Müller, O., Seuthe, L., Bratbak, G., and Paulsen, M. L. (2018). Bacterial response to permafrost derived organic matter input in an arctic fjord. *Front. Mar. Sci.* 5:263. doi: 10.3389/fmars.2018.00263

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

Nelson, C. E., and Carlson, C. A. (2012). Tracking differential incorporation of dissolved organic carbon types among diverse lineages of Sargasso Sea bacterioplankton. *Environ. Microbiol.* 14, 1500–1516. doi: 10.1111/j.1462-2920.2012.02738.x

Nelson, C. E., and Wear, E. K. (2014). Microbial diversity and the lability of dissolved organic carbon. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7166–7167. doi: 10.1073/pnas.1405751111

Obu, J., Lantuit, H., Grosse, G., Günther, F., Sachs, T., Helm, V., et al. (2016). Coastal erosion and mass wasting along the Canadian Beaufort Sea based on annual airborne LiDAR elevation data. *Geomorphology* 293, 331–346. doi: 10.1016/j.geomorph.2016.02.014

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., et al. (2019). Northern Hemisphere permafrost map based on TTOP modelling for 2000-2016 at 1 km2 scale. *Earth Sci. Rev.* 193, 299–316. doi: 10.1016/j.earscirev.2019.04.023

O'Donnell, J. A., Aiken, G. R., Swanson, D. K., Panda, S., Butler, K. D., and Baltensperger, A. P. (2016). Dissolved organic matter composition of Arctic rivers: linking permafrost and parent material to riverine carbon. *Glob. Biogeochem. Cycles* 30, 1811–1826. doi: 10.1002/2016GB005482

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019). *vegan: Community Ecology Package*. Available online at: https://cran.r-project.org/web/packages/vegan/index.html (accessed May 28, 2020).

Opfergelt, S. (2020). The next generation of climate model should account for the evolution of mineral-organic interactions with permafrost thaw. *Environ. Res. Lett.* 15:091003. doi: 10.1088/1748-9326/ab9a6d

Overeem, I., Anderson, R. S., Wobus, C. W., Clow, G. D., Urban, F. E., and Matell, N. (2011). Sea ice loss enhances wave action atthe Arctic coast. *Geophys. Res. Lett.* 38:L17503. doi: 10.1029/2011GL0 48681

Palmtag, J., and Kuhry, P. (2018). Grain size controls on cryoturbation and soil organic carbon density in permafrost-affected soils. *Permafrost Periglac. Process.* 29, 112–120. doi: 10.1002/ppp.1975

Park, S., Kim, I. K., Lee, J.-S., and Yoon, J.-H. (2019). *Sulfitobacter sabulilitoris* sp. nov., isolated from marine sand. *Int. J. Syst. Evol. Microbiol.* 69, 3230–3236. doi: 10.1099/ijsem.0.003614

Paulsen, M. L., Nielsen, S. E. B., Müller, O., Møller, E. F., Stedmon, C. A., Juul-Pedersen, T., et al. (2017). Carbon bioavailability in a high arctic fjord influenced by glacial meltwater, NE Greenland. *Front. Mar. Sci.* 4:176. doi: 10.3389/fmars.2017.00176

Pollard, W. (2018). "Periglacial processes in glacial environments," in *Past Glacial Environments*, eds J. Menzies and J. J. M. van der Mee (Amsterdam: Elsevier), 537–564. doi: 10.1016/B978-0-08-100524-8.00016-6

Pomeroy, L. R., and Wiebe, W. J. (2001). Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquat. Microb. Ecol.* 23, 187–204. doi: 10.3354/ame023187

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rampton, V. N. (1982). *Quaternarygeology of the Yukon Coastal Plain*. Ottawa: Geological Survey of Canada.

Raudina, T. V., Loiko, S. V., Lim, A. G., Krickov, I. V., Shirokova, L. S., Istigechev, G. I., et al. (2017). Dissolved organic carbon and major and trace elements in peat porewater of sporadic, discontinuous, and continuous permafrost zones of western Siberia. *Biogeosciences* 14, 3561–3584. doi: 10.5194/bg-14-3561-2017

Robinson, M. D., and Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9, 321–332. doi: 10.1093/biostatistics/kxm030

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Romanenko, L. A., Tanaka, N., Svetashev, V. I., and Kalinovskaya, N. I. (2011). *Pacificibacter maritimus* gen. nov., sp. nov., isolated from shallow marine sediment. *Int. J. Syst. Evol. Microbiol.* 61, 1375–1381. doi: 10.1099/ijs.0.026047-0

Romanovsky, V. E., Smith, S. L., and Christiansen, H. H. (2010). Permafrost thermal state in the polar Northern Hemisphere during the international polar year 2007-2009: a synthesis. *Permafrost Periglac. Process.* 21, 106–116. doi: 10.1002/ppp.689

Rosenstock, B., and Simon, M. (2001). Sources and sinks of dissolved free amino acids and protein in a large and deep mesotrophic lake. *Limnol. Oceanogr.* 46, 644–654. doi: 10.4x319/lo.2001.46.3.0644

Ruiz-González, C., Niño-García, J. P., and Del Giorgio, P. A. (2015). Terrestrial origin of bacterial communities in complex boreal freshwater networks. *Ecol. Lett.* 18, 1198–1206. doi: 10.1111/ele.12499

Schirrmeister, L., Grosse, G., Wetterich, S., Overduin, P. P., Strauss, J., Schuur, E. A. G., et al. (2011). Fossil organic matter characteristics in permafrost deposits of the northeast Siberian Arctic. *J. Geophys. Res.* 116:G00M02. doi: 10.1029/2011JG001647

Schnetger, B., and Lehners, C. (2014). Determination of nitrate plus nitrite in small volume marine water samples using vanadium(III)chloride as a reductionagent. *Mar. Chem.* 160, 91–98. doi: 10.1016/j.marchem.2014.01.010

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., et al. (2015). Climate change and the permafrost carbon feedback. *Nature* 520, 171–179. doi: 10.1038/nature14338

Semiletov, I. P., Shakhova, N. E., Pipko, I. I., Pugach, S. P., Charkin, A. N., Dudarev, O. V., et al. (2013). Space-time dynamics of carbon and environmental parameters related to carbon dioxide emissions in the Buor-Khaya Bay and adjacent part of the Laptev Sea. *Biogeosciences* 10, 5977–5996. doi: 10.5194/bg-10-5977-2013

Sipler, R. E., Kellogg, C. T. E., Connelly, T. L., Roberts, Q. N., Yager, P. L., and Bronk, D. A. (2017). Microbial community response to terrestrially derived dissolved organic matter in the coastal arctic. *Front. Microbiol.* 8:1018. doi: 10.3389/fmicb.2017.01018

Sjöstedt, J., Hagström, Å, and Zweifel, U. L. (2012a). Variation in cell volume and community composition of bacteria in response to temperature. *Aquat. Microb. Ecol.* 66, 237–246. doi: 10.3354/ame01579

Sjöstedt, J., Koch-Schmidt, P., Pontarp, M., Canbäck, B., Tunlid, A., Lundberg, P., et al. (2012b). Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Appl. Environ. Microbiol.* 78, 1361–1369. doi: 10.1128/AEM.05542-11

Smith, S. L., Romanovsky, V. E., Lewkowicz, A. G., Burn, C. R., Allard, M., Clow, G. D., et al. (2010). Thermal state of permafrost in North America: a contribution to the international polar year. *Permafrost Periglac. Process.* 21, 117–135. doi: 10.1002/ppp.690

Sorokin, D. Y. (1995). *Sulfitobacter pontiacus* gen. nov., sp. nov. — A new heterotrophic bacterium from the black sea, specialized on sulfite oxidation. *Mikrobiologiya* 64:295.

Spencer, R. G. M., Mann, P. J., Dittmar, T., Eglinton, T. I., McIntyre, C., Holmes, R. M., et al. (2015). Detecting the signature of permafrost thaw in Arctic rivers. *Geophys. Res. Lett.* 42, 2830–2835. doi: 10.1002/2015GL063498

Stapel, J. G., Schwamborn, G., Schirrmeister, L., Horsfield, B., and Mangelsdorf, K. (2017). Substrate potential of last interglacial to Holocene permafrost organic matter for future microbial greenhouse gas production. *Biogeosciences* 15, 1969–1985. doi: 10.5194/bg-15-1969-2018

Stedmon, C. A., and Nelson, N. B. (2015). "The optical properties of DOM in the ocean," in *Biogeochemistry of Marine Dissolved Organic Matter*, eds D. A. Hansell and C. A. Carlson (Amsterdam: Elsevier), 481–508. doi: 10.1016/B978- 0-12-405940-5.00010-8

Strauss, J., Schirrmeister, L., Grosse, G., Fortier, D., Hugelius, G., Knoblauch, C., et al. (2017). Deep Yedoma permafrost: a synthesis of depositional characteristics and carbon vulnerability. *Earth Sci. Rev.* 172, 75–86. doi: 10. 1016/j.earscirev.2017.07.007

Svoboda, P., Lindström, E. S., Ahmed Osman, O., and Langenheder, S. (2018). Dispersal timing determines the importance of priority effects in bacterial communities. *ISME J.* 12, 644–646. doi: 10.1038/ismej.2017.180

Tank, S. E., Vonk, J. E., Walvoord, M. A., McClelland, J. W., Laurion, I., and Abbott, B. W. (2020). Landscape matters: predicting the biogeochemical effects of permafrost thaw on aquatic networks with a state factor approach. *Permafrost Periglac. Process.* 31, 358–370. doi: 10.1002/ppp.2057

Tanski, G., Couture, N., Lantuit, H., Eulenburg, A., and Fritz, M. (2016). Eroding permafrost coasts release low amounts of dissolved organic carbon (DOC) from ground ice into the nearshore zone of the Arctic Ocean. *Glob. Biogeochem. Cycles* 30, 1054–1068. doi: 10.1002/2015GB005337

Tanski, G., Wagner, D., Knoblauch, C., Fritz, M., Sachs, T., and Lantuit, H. (2019). Rapid co_2 release from eroding permafrost in seawater. *Geophys. Res. Lett.* 46, 11244–11252. doi: 10.1029/2019GL084303

Textor, S. R., Guillemette, F., Zito, P. A., and Spencer, R. G. M. (2018). An assessment of dissolved organic carbon biodegradability and priming in blackwater systems. *J. Geophys. Res. Biogeosci.* 123, 2998–3015. doi: 10.1029/2018JG004470

Textor, S. R., Wickland, K. P., Podgorski, D. C., Johnston, S. E., and Spencer, R. G. M. (2019). Dissolved organic carbon turnover in permafrost-influenced watersheds of interior alaska: molecular insights and the priming effect. *Front. Earth Sci.* 7:275. doi: 10.3389/feart.2019.00275

Thingstad, T. F., Bellerby, R. G. J., Bratbak, G., Børsheim, K. Y., Egge, J. K., Heldal, M., et al. (2008). Counterintuitive carbon-to-nutrient coupling in an Arctic pelagic ecosystem. *Nature* 455, 387–390. doi: 10.1038/nature07235

Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M., and Michel, G. (2011). Environmental and gut bacteroidetes: the food connection. *Front. Microbiol.* 2:93. doi: 10.3389/fmicb.2011.00093

Van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., et al. (2007). The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20404–20409. doi: 10.1073/pnas.07072 00104

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., and Cushing, C. E. (1980). The river continuum concept. *Can. J. Fish. Aquat. Sci.* 37, 130–137. doi: 10.1139/f80-017

Vonk, J. E., Mann, P. J., Davydov, S., Davydova, A., Spencer, R. G. M., Schade, J., et al. (2013). High biolability of ancient permafrost carbon upon thaw. *Geophys. Res. Lett.* 40, 2689–2693. doi: 10.1002/grl.50348

Vonk, J. E., Sánchez-García, L., van Dongen, B. E., Alling, V., Kosmach, D., Charkin, A., et al. (2012). Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia. *Nature* 489, 137–140. doi: 10.1038/nature11392

Vonk, J. E., Tank, S. E., Bowden, W. B., Laurion, I., Vincent, W. F., Alekseychik, P., et al. (2015). Reviews and syntheses: effects of permafrost thaw on Arctic aquatic ecosystems. *Biogeosciences* 12, 7129–7167. doi: 10.5194/bg-12-7129-2015

Walker, S. A., Amon, R. M. W., and Stedmon, C. A. (2013). Variations inhigh-latitude riverine fluorescent dissolved organic matter: a comparison of large Arctic rivers. *J. Geophys. Res. Biogeosci.* 118, 1689–1702. doi: 10.1002/2013JG002320

Walter Anthony, K., Schneider von Deimling, T., Nitze, I., Frolking, S., Emond, A., Daanen, R., et al. (2018). 21st-century modeled permafrost carbon emissions accelerated by abrupt thaw beneath lakes. *Nat. Commun.* 9:3262. doi: 10.1038/s41467-018-05738-9

Wang, Y., Naumann, U., Wright, S. T., and Warton, D. I. (2012). mvabund - anR package for model-based analysis of multivariate abundance data. *Methods Ecol. Evol.* 3, 471–474. doi: 10.1111/j.2041-210X.2012.00190.x

Ward, C. P., and Cory, R. M. (2015). Chemical composition of dissolved organic matter draining permafrost soils. *Geochim. Cosmochim. Acta* 167, 63–79. doi: 10.1016/j.gca.2015.07.001

Weishaar, J. L., Aiken, G. R., Bergamaschi, B. A., Fram, M. S., Fujii, R., and Mopper, K. (2003). Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environ. Sci. Technol.* 37, 4702–4708. doi: 10.1021/es030360x

Wickland, K. P., Waldrop, M. P., Aiken, G. R., Koch, J. C., Jorgenson, M. T., and Striegl, R. G. (2018). Dissolved organic carbon and nitrogen release from boreal Holocene permafrost and seasonally frozen soils of Alaska. *Environ. Res. Lett.* 13:065011. doi: 10.1088/1748-9326/aac4ad

Wild, B., Andersson, A., Bröder, L., Vonk, J., Hugelius, G., McClelland, J. W., et al. (2019). Rivers across the Siberian Arctic unearth the patterns of carbon release from thawing permafrost. *Proc. Natl. Acad. Sci. U.S.A.* 116, 10280–10285. doi: 10.1073/pnas.1811797116

Wologo, E., Shakil, S., Zolkos, S., Textor, S., Ewing, S., Klassen, J., et al. (2020). Stream dissolved organic matter in permafrost regions shows surprising compositional similarities but negative priming and nutrient effects. *Glob. Biogeochem. Cycles* 35:e2020GB006719. doi: 10.1029/2020GB006719

Wolter, J., Lantuit, H., Herzschuh, U., Stettner, S., and Fritz, M. (2017). Tundra vegetation stability versus lake-basin variability on the Yukon Coastal Plain (NW Canada) during the past three centuries. *Holocene* 27, 1846–1858. doi: 10.1177/0959683617708441

Yamashita, Y., and Tanoue, E. (2003). Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. *Mar. Chem.* 82, 255–271. doi: 10.1016/S0304-4203(03)00073-2

Yoon, J.-H., Kang, S.-J., and Oh, T.-K. (2005). *Marinomonas dokdonensis* sp. nov., isolated from sea water. *Int. J. Syst. Evol. Microbiol.* 55, 2303–2307. doi: 10.1099/ijs.0.63830-0

Zhang, D.-C., Li, H.-R., Xin, Y.-H., Liu, H.-C., Chen, B., Chi, Z.-M., et al. (2008). *Marinomonas arctica* sp. nov., a psychrotolerant bacterium isolated from the Arctic. *Int. J. Syst. Evol. Microbiol.* 58, 1715–1718. doi: 10.1099/ijs.0.65737-0

Zhang, Z., Qin, J., Sun, H., Yang, J., and Liu, Y. (2020). Spatiotemporal dynamics of dissolved organic carbon and freshwater browning in the Zoige Alpine Wetland, Northeastern Qinghai-Tibetan Plateau. *Water* 12:2453. doi: 10.3390/w12092453

Zweifel, U. L., Blackburn, N., and Hagström, Å (1996). Cycling of marine dissolved organic matter. I. An experimental system. *Aquat. Microb. Ecol.* 11, 65–77. doi: 10.3354/ame011065

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Paper B

Lignin phenol quantification from machine learning-assisted liquid chromatography-absorbance spectroscopy data.

Anders Dalhoff Bruhn, Urban Wünsch, Christopher L. Osburn, Jacob C. Rudolph and Colin A. Stedmon.

This paper is in second review at Limonology & Oceanography: Methods.

Lignin phenol quantification from machine learning-assisted decomposition of liquid chromatography-absorbance spectroscopy data.

Anders Dalhoff Bruhn[1], Urban Wünsch[1], Christopher L. Osburn[2], Jacob C. Rudolph[2,3] and Colin A. Stedmon[1].

[1]National Institute of Aquatic Resources, Technical University of Denmark, Denmark.
[2]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, USA.
[3]Department of Environmental Biology, Smithsonian Environmental Research Center, Edgewater, MD

Correspondence to: Anders Dalhoff Bruhn (adbj@aqua.dtu.dk) and Colin A. Stedmon (cost@aqua.dtu.dk)

Running head: Machine learning-assisted lignin phenol quantification

Keywords: Lignin, spectroscopy, absorbance, seawater, HPLC, PARAFAC2

Abstract

Analysis of lignin in seawater is essential to understanding the fate of terrestrial dissolved organic matter (DOM) in the ocean and its role in the carbon cycle. Lignin is typically quantified by gas or liquid chromatography, coupled with mass spectrometry (GC-MS or LC-MS). MS instrumentation can be relatively expensive to purchase and maintain. Here we present an improved approach for quantification of lignin phenols using LC and absorbance detection. The approach applies a modified parallel factor analysis algorithm (PARAFAC2) to 2nd derivative absorbance chromatograms. It is capable of isolating individual elution profiles of analytes despite co-elution and overall improves sensitivity and specificity, compared to manual integration methods. For most lignin phenols, detection limits below 5 nM were achieved, which is comparable to MS detection. The reproducibility across all laboratory stages for our reference material showed a relative standard deviation between 1.47-16.84% for all 11 lignin phenols. Changing the amount of DOM in the reaction vessel for the oxidation (dissolved organic carbon between 22-367 mM), did not significantly affect the final lignin phenol composition. The new method was applied to seawater samples from the Kattegat and Davis Strait. The total concentration of dissolved lignin phenols measured in the two areas was between 4.3 - 10.09 and 2.1 - 3.2 nM, respectively, which is within the range found by other studies. Comparison with a different oxidation approach and detection method (GC-MS) gave similar results and further underline the potential of LC and absorbance detection for lignin analysis in natural water with our proposed method.

Introduction

Anthropogenic activities and climate change have increased the mobilization of terrestrial organic matter (plant detritus, humic substances, and other organic compounds deriving from land) from soils into rivers, lakes and eventually the ocean (Freeman et al., 2001; Gardiner & Miller, 2004; Quinton et al., 2010; Regnier et al., 2013) and accelerated permafrost thaw has led to additional release of ancient terrestrial organic matter, mainly derived from plant detritus, into aquatic environments (Spencer et al., 2015; Vonk et al., 2012, 2013; Wild et al., 2019). The fate of terrestrial dissolved organic matter (tDOM) is important to determine with respect to understanding its role in a changing carbon cycle (Ciais et al., 2008; Cole et al., 2007). Concentrations of dissolved lignin – a biochemical marker for tDOM in seawater – are generally low. For example, in the Pacific Ocean tDOM on average makes up around 1% of the total dissolved organic carbon (DOC) (Hernes & Benner, 2002). However, in the Arctic Ocean surface waters 14-24% of the DOC is found to be terrestrial (Benner et al., 2005), which reflects the closer proximity to riverine inputs from surrounding continents in the Arctic, compared to that of the large open ocean basins.

Lignin is an amorphous and highly branched phenolic biopolymer used as a building block in plant cell walls to create structural support (Monties & Fukushima, 2001). Lignin exists only in vascular plants (Lewis & Yamamoto, 1990; Monties & Fukushima, 2001) and is therefore an excellent biomarker of terrestrial plant material. For measurement, the lignin macrostructure must first be oxidized into its constituent phenolic monomers, which – in addition to the phenol functionality – may carry aldehyde-, ketone- and acid functional groups. These lignin-derived phenols can also be divided into groups depending on ring substitution (p-hydroxyl (P), vanillyl (V), syringyl (S)) and conjugation (e.g. cinnamyl (C)). The lignin concentration in DOM is reported as the total dissolved lignin phenols (TDLP) measured after its oxidation. Open ocean TDLP concentrations are at pM level, while the coastal ocean ranges values up to 500 nM, depending on proximity to river discharge (Fichot et al., 2016). In estuaries, lignin is often strongly correlated with DOC, as both have a high concentration source in freshwater, and mix more or less conservatively with salinity (Fichot et al., 2016; Hernes & Benner, 2003; Osburn et al., 2016). Ratios between P/V, C/V and S/V can be used to determine the source of terrestrial material (Amon et al., 2003; Hedges & Mann, 1979; Lobbes et al., 2000; Mann et al., 2016) and to reflect diagenetic fate in aquatic environments (Hernes & Benner, 2003; Kaiser et al., 2017; Opsahl & Benner, 1998).(Hernes & Benner, 2003; Kaiser et al., 2017; Opsahl & Benner, 1998).

Quantification of TDLP in seawater requires a sequence of steps including filtration, solid phase extraction (SPE), oxidation, purification, and finally quantification of individual lignin phenols. Filtration removes particulate matter and subsequent SPE concentrates the DOM (including dissolved lignin) and desalts the sample prior to oxidation (Dittmar et al., 2008). The oxidation is performed in a reaction vessel (typically a sealed Teflon or metal cylinder) and is mediated by the addition of a copper oxidant (Hedges & Ertel, 1982; Yan & Kaiser, 2018a). The reaction products are isolated from the reaction mixture afterwards using a hydrophilic lipophilic balanced (HLB) cartridge (Yan & Kaiser, 2018b) containing a resin made from a co-polymer of divinybenzene and N-vinylpyrrolidone. The subsequently eluted lignin phenols are quantified using a combination of high-performance liquid chromatography (HPLC) or gas chromatography (GC), coupled with absorbance spectroscopy or MS (Hedges & Ertel, 1982; Kaiser & Benner, 2012a; Lobbes et al., 1999; Louchouarn et al., 2010; Yan & Kaiser, 2018b). The downscaling of the reaction vessel size (thereby volumes of reactants needed) together with the alternative use of $CuSO_4$ as oxidant, has improved yield of lignin phenols, and offers superior accuracy and precision (Yan & Kaiser, 2018a).

Due to its high sensitivity and specificity, MS has been the preferential quantification method for lignin phenols. However, widespread application of MS is limited by the high purchase and maintenance costs of instrumentation. In comparison, HPLC-absorbance spectroscopy traditionally struggles with lower sensitivity, lower specificity and increased background interference from DOM compared to MS. The resulting lower sensitivity therefore requires the use of more water for extraction and makes the HPLC-absorbance spectroscopy approach less desirable.

Machine-learning approaches have become more readily accessible, and a modified version of parallel factor analysis (PARAFAC2) has shown great potential to overcome problems with shifting, overlapping, and low intensity peaks in chromatography (Amigo et al., 2008, 2010). Traditional PARAFAC finds unique underlying solutions of independent components (analytes) and scores (concentrations) across a range of samples and has, among other applications, been widely used to characterize excitation-emission fluorescence matrices of DOM samples (Bro, 1997; Murphy et al., 2013; Stedmon et al., 2003). In contrast to PARAFAC, PARAFAC2 is less constrained and allows one of the modes to shift to a minor extent in between samples, which makes it more suitable for chromatographic data where shifts in retention time and changes of peak shape often occur (Bro et al., 1999; Harshman, 1972).

This study is therefore focused on the development of machine-learning assisted absorbance-based HPLC approach, which mitigates current limitations (with sensitivity and specificity) and allows the quantification of lower concentrations of lignin phenols in a complex mixture of DOM and hence require lower initial sample volumes of seawater.

Materials & procedures

Sampling

All plastic material used in the study was acid-cleaned and ultrapure water rinsed before use, whereas all glassware was acid-cleaned and combusted. Seawater samples for analysis were taken at entrance to the Baltic (Kattegat) and in the Davis Strait west of Baffin Island, Canada (see Table 1 for sampling details). Both cruises were conducted with the research vessel R/V Dana during autumn 2021.

Seawater samples were taken using a Niskin water sampler mounted on a CTD rosette (Seabird Scientific). Between 2.4 and 4.8 L of seawater was collected for each sample and filtered through a 0.2 µm cartridge filter (Polyethersulfone Membrane Capsule Filter, Sterlitech Inc.) using a peristaltic pump. After filtration the samples were acidified to approximately pH 2 by addition of hydrochloric acid (HCl). Subsamples (60 mL) were collected from the acidified seawater samples in combusted brown glass vials for measurement of DOC. The seawater samples were then stored dark and cold (5 °C) until extraction.

Table 1: Location, time, depth and volume extracted of samples collected for testing the method. Also shown are the salinity and dissolved organic carbon concentrations.

Location	Date	Coordinate	Station	Depth (m)	Volume extracted (L)	Salinity (‰)	DOC (µM)
Davis Strait	10-09-2020	66°41’54" N 60°46'47" W	3	100	4.80	32.8	65
Davis Strait	10-09-2020	66°43'49” N 60°29'57" W	4	100	4.80	32.9	68
Davis Strait	12-09-2020	66°44'57” N 60°30'48" W	5	100	4.80	32.8	72
Danish waters	26-10-2020	55°38’09” N 10°42’24” E	6	5	2.45	20.9	ND
Danish waters	26-10-2020	55°38’09” N 10°42’24” E	6	15	2.40	25.5	227
Danish waters	26-10-2020	55°38’09” N 10°42’24” E	6	20	2.40	28.1	157
Danish waters	28-10-2020	55°55’54” N 12°37’87” E	11	4	2.45	18.9	ND
Danish waters	28-10-2020	55°55’54” N 12°37’87” E	11	15	2.40	31.8	150
Danish waters	28-10-2020	55°55’54” N 12°37’87” E	11	19	2.40	32.8	113
Danish waters	28-10-2020	56°41’17” N 11°44’24” E	12	4	2.40	21.6	180
Danish waters	28-10-2020	56°41’17” N 11°44’24” E	12	13	2.40	25.0	165
Danish waters	28-10-2020	56°41’17” N 11°44’24” E	12	31	2.40	33.4	103

Solid Phase Extraction

Solid phase extraction of DOM was performed on the filtered and acidified seawater using Bond Elut Priority PolLutant (PPL) cartridges (200 mg sorbent, 3 mL cartridge capacity, Agilent) according to the method by Dittmar et al. (2008) with minor modifications. The samples were introduced into the PPL cartridges using PTFE tubing and tube adapters (Supelco) at a flowrate between 4-6 mL/min using a compact precision peristaltic pump (Shenchen). Before use, the PPL cartridges were cleaned with 3 mL methanol and conditioned with 6 mL of acidified ultrapure water (pH 2, HCl). After the extraction of DOM from the seawater samples, the sorbent was rinsed with 6 mL of acidified ultrapure water (pH 2, HCl) to remove salts. The sorbent was dried with a vacuum manifold using a pressure of – 60 kPa for 5 minutes. The cartridges were then stored in acid cleaned and combusted brown vials until further processing. The volume of the extracted seawater was determined using a 1 L measuring cylinder (± 10 mL) to be able to calculate the enrichment factor and approximate the environmental concentration of lignin phenols. From the extracted seawater, 60 mL of water was collected in brown glass vials for DOC measurement.

Dissolved organic carbon

Dissolved organic carbon was determined using high-temperature catalytic combustion (TOC/VCPH, Shimadzu). Fifteen mL of acidified sample (collected before SPE extraction, pH 2 by HCl) was poured into a clean glass vial (550°C, 5 h). Samples were sparged in the autosampler for the instrument using oxygen gas to remove all inorganic carbon prior to injection. A 100 µL sample per determination was injected onto the catalyst and a minimum of three injections were averaged to determine the mean instrument response. The detector response was converted to DOC concentrations via a seven-point standard curve between 0-311 µM carbon using acetanilide. To ensure calibration stability, ultrapure water spiked with 62 µM and 104 µM standards served as reference samples, which were all repeatedly measured after every 7th sample. Additionally, the instrument performance was verified by determining that the DOC concentrations of the deep-sea community reference standard (Hansell laboratory, University of Miami) fell between 42 and 45 µM.

Cupric oxidation

The method for cupric oxidation applied in our study was a modified version of that developed by Yan & Kaiser (2018a). In-house built stainless steel reaction vessels with a volume of 850 µL were thoroughly cleaned before use by soaking in NaOH, then rinsed in methanol, soaked in ethanol overnight and finally rinsed multiple times with ultrapure water. In between samples, the cylinders and threads were submerged in ethanol for half an hour and rinsed thoroughly with ultrapure water afterwards.

The DOM retained on the PPL cartridge was eluted with 3 mL methanol (HPLC grade, Sigma Aldrich). Afterwards the eluate was evaporated under nitrogen gas and redissolved in 3 mL of methanol for a second time (to achieve exact volumes). From this 3 mL methanol, either the whole volume (for 7 samples) or 0.6 mL (for 5 samples) was transferred to the reaction vessel. The methanol was then evaporated in a fume hood with a gentle stream of nitrogen gas. Afterwards 45 µL of 10 mM $CuSO_4$, 40 µL of 0.2 M ascorbic acid (antioxidant to avoid over-oxidation), and 748 µL of 1.1 M NaOH were pipetted into the reaction vessel containing the dried DOM. Adding these volumes of reactant to the reaction vessel led to almost identical concentrations of each compared to those from Yan & Kaiser (2018a). The total pipetted volume filled the whole volume of the reaction vessel, leaving no headspace. The sealed reaction vessel were placed in an oven at 150 °C for two hours, which was found to give the highest yield of lignin phenols according to Yan & Kaiser (2018a). Afterwards the vessel was cooled quickly using water and the reaction product was transferred into a 5 mL combusted glass vial. The vessel was rinsed with 3 x 800 µL of ultrapure water to make sure all of the mixture was transferred to the 5 mL glass vial. The dilution of the reaction vessel products with ultrapure water also avoided flocculation of the DOM upon subsequent acidification (data not shown).

HLB extraction and clean-up

The oxidized organics from the reaction vessel were purified using a HLB cartridge as described in Yan & Kaiser (2018b). Prior to the HLB extraction, 17 µL of 0.25 mM cinnamic acid (CIN) was added (to give a final concentration of 21.25 µM) as an internal standard to account for loss of lignin phenols during the HLB extraction and the subsequently clean-up steps. After addition of CIN, the solution was acidified to pH 2 with 85 µL of 6 M H_2SO_4. The cartridge was placed in the vacuum manifold and cleaned and conditioned with 2 x 1 mL

methanol and 2 x 1 mL of acidified ultrapure water (pH~2, 7.4 mM H_3PO_4). A vacuum hand pump was used to maintain a low flow (pressure at -25 kPa) across the HLB cartridge. The acidified cupric oxidized product solution (approximately 3.2 mL) was pipetted onto the HLB cartridge and subsequently 3 x 300 µL of a 20/80 (v/v%) methanol/water mixture was applied to remove inorganics and weakly retained compounds (Yan & Kaiser, 2018b). Finally, nitrogen gas was applied for 5 minutes to dry the sorbent. Elution of the sample into a 1.5 mL glass vial was carried out with a 30/70 (v/v%) methanol/methyl acetate mixture. The addition of methanol to methyl acetate assists with the elution of acidic phenols that are strongly bound to the resin due to hydrogen bridge interactions (Yan & Kaiser, 2018b). To remove the residual methanol/methyl acetate from the cartridge, nitrogen gas was briefly applied. The methanol/methyl acetate was evaporated using nitrogen gas, and the sample was reconstituted in 190 µL ultrapure water with pH adjusted to 2.5 with 10 µL of 0.14 M H_3PO_4. The 200 µL aqueous sample containing the oxidized DOM was then transferred to a HPLC vial with a 250 µL insert vial and stored at -18 °C until analysis.

Analytical standards

Pure lignin phenol monomers were used as analytical calibration standards. The 11 natural lignin phenols are named and abbreviated systemically, first by their chemical group related to the amount of methoxy groups on the benzene structure, i.e. p-hydroxyl (P), syringyl (S), vanillyl (V), coumaric (C as in CAD) and ferulic (F), followed by the type of functional group, additional to the phenol, i.e. acid (AD), aldehyde (AL) and ketone (ON). The internal standard (CIN) was not named according to this system.

P-hydroxybenzoic acid (PAD), coumaric acid (CAD) and CIN were obtained from Merck. P-hydroxybenzaldehyde (PAL), syringic acid (SAD), vanillin (VAN) and syringaldehyde (SAL) were obtained from Alfa Aesar. Acetovanillone (VON), ferulic acid (FAD) and acetosyringone (SON) were obtained from Acros Organics. P-hydroxyacetophenone (PON) and vanillic acid (VAD) were obtained from Sigma Aldrich and TCI Europe, respectively. Stock solutions with a concentration of 1.25 mM in ultrapure water were made for each of the 11 lignin phenols and the internal standard. The solution was stirred with a magnetic stirrer over the course of two days to make sure complete dissolution was achieved. The stock solutions were afterwards stored at -18 °C until further use.

Calibration curves consisted of 13 concentrations between 1 nM - 25 µM and all phenols were combined to yield standard mixtures containing different concentrations of each compound (calibration standards 1-13, Table S1). This approach ensured that any effect of co-elution between phenols with concentration differences (e.g. 25 µM and 10 µM) could be addressed when applying the 2nd derivative/PARAFAC2 method to create calibration curves. To calculate limit of detection (LOD), additional calibration standards were made with concentrations mixed between 1 and 25 nM (calibration standards 14-18, Table S1). A standard mixture (20 µM of each lignin phenol) was made by diluting the stock solutions in ultrapure water. To evaluate the recovery of lignin phenols and the internal standard in the presence of DOM background matrix, a selected seawater sample (SE Kattegat, st. 11/19m) was spiked with the standard mixture so it contained known added concentrations (~5 µM) of each lignin phenol. Both the standard mixture and the spiked sample was injected between every 7th sample.

To assess the reproducibility of the laboratory and analytical procedures, a solution was made from Suwannee River natural organic matter reference material (SRNOM, Cat. Num. 2R101N). The SRNOM was purchased from the International Humic Substances Society, isolated in 2012 using reverse osmosis (Green et al., 2015). 150 mg of solid SRNOM was dissolved in 25 mL of methanol to yield approx. 250 mM C. Three replicate samples, made from the SRNOM stock solution, were then subjected to cupric oxidation, clean up and analysis. Adding 120 µL of the SRNOM stock solution to the reaction vessel (the stainless-steel cylinder for cupric oxidation in the oven), this resulted in the addition of approximately 30 µmol C, which is similar to that extracted from the natural seawater samples.

Chromatography and detection

High-pressure liquid chromatography was performed on a Nexera X2 HPLC system (Shidmadzu) equipped with LC-20AB pumps using a C18 column (Poroshell 120 EC-C18 4.6 x 100 mm, 2.7 µm particle diameter, Agilent) with a guard column attached. The C18 column and the guard column were heated to 50 °C using a column oven. The two mobile phases used for HPLC consisted of 7.4 mM phosphoric acid in ultrapure water (mobile phase A; pH 2.4) and pure acetonitrile (mobile phase B). Many studies use a mixture of acetonitrile and methanol to elute phenolic compounds (Fischer & Höffler, 2021; Ingalls et al., 2010; Lobbes et al., 1999; Steinberg et al., 1984). Here, we elected to use pure acetonitrile to speed up the elution and obtain compound absorbance spectra with less solvent interference. The total runtime for each

sample was 40 minutes. Between 0-18 minutes, mobile phase B increases from 5% to 55% (elution of lignin phenols and the internal standard), between 18-20 minutes mobile B increases to 100% and afterwards stays at 100% until the 22 minute mark (removing all residual organic compounds from the column). Between 22-35 minutes mobile phase B decreases back to 5% and stays there until the 40 minute mark to stabilize the pressure before the next run starts. The mobile phases were pumped through the columns at a constant flowrate of 1 mL/min and injection volume was kept at 50 µL for all samples and standards. The diode array detector (DAD; SPD-M30A, Shimadzu) measured absorbance between 240 and 700 nm at a rate of 1.5 Hz.

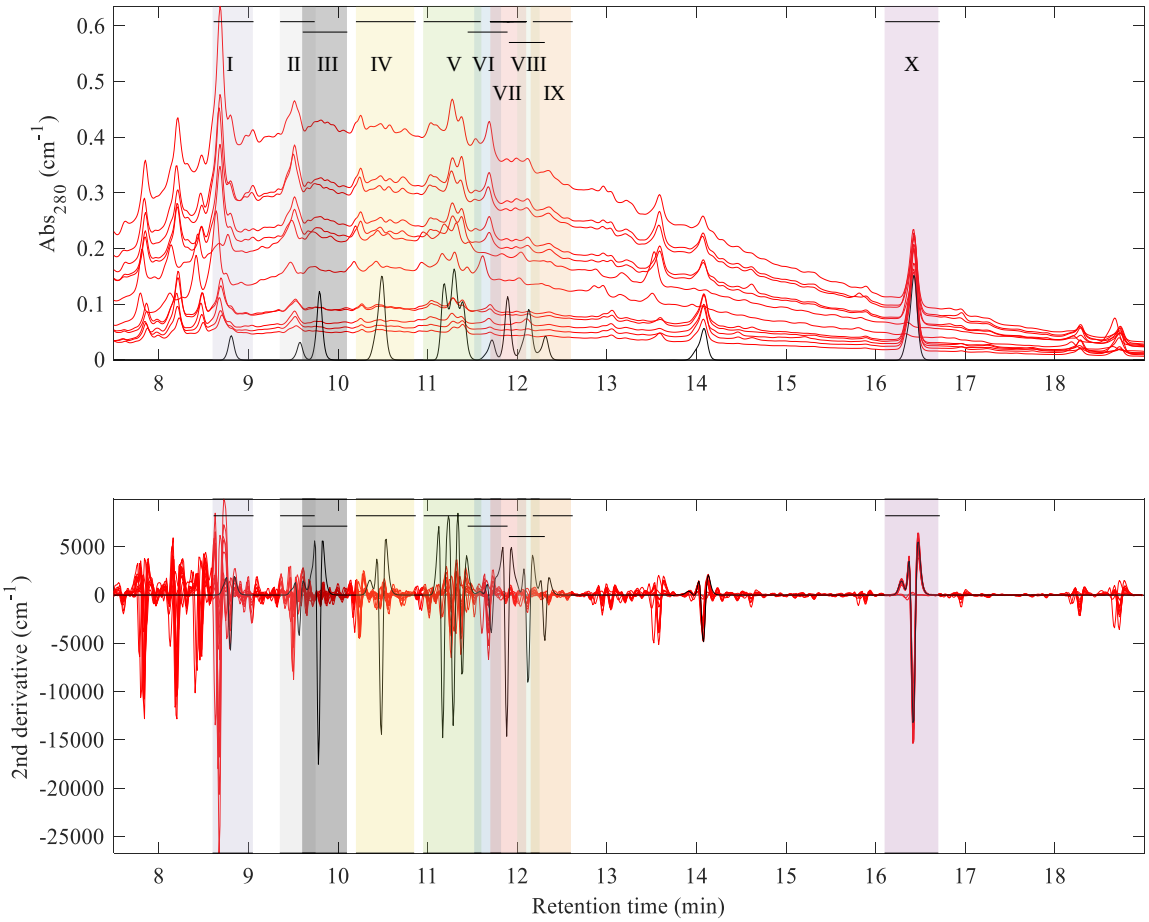


Figure 1: Raw chromatograms (a) and 2nd derivative chromatograms (b) of a standard mixture (black line) and natural samples (red lines) at 280 nm absorbance. Intervals indicate where lignin phenols are eluting and is numbered with numerals from I to X. The chromatograms have been corrected so that the retention time of CIN peaks (interval X) align across samples. Table 3 lists which lignin phenols elute in a certain interval. The upper black lines indicate the limits of the intervals (which in most cases overlap).

PARAFAC2 modelling

The isolation of individual elution profiles and spectra of the lignin phenols was carried out using PARAFAC2. The analysis fits the following equation to the chromatographic data using an alternating least square routine:

$$x_{ijk} = \sum_{f=1}^{F} a_{if}^{k} * b_{jf} * c_{kf} + e_{ijk} \qquad \text{eq. (I)}$$

$$k = 1,2,\dots, K\ ;\ j = 1,2,\dots, J\ ;\ i = 1,2,\dots, I$$

Where x_{ijk} corresponds to the data points in the chromatographic output (sample x retention time x wavelength) from the HPLC-DAD instrument. For x_{ijk}, *k* corresponds to sample, *i* corresponds to retention time, and *j* corresponds to absorbance wavelength. Each PARAFAC2 component *f* is described by three vectors: the concentration, a^k_{if}, the elution profile (retention time), b_{jf}, and the absorbance spectrum, c_{kf}. Each element in x_{ijk}, is calculated as the sum of abundance recorded across the predefined number of components, *F*. The superscript *k* in a^k_{if} implies that the elution profiles in between samples can deviate slightly from each other and allow for minor retention time shifts between samples. The residual (unexplained) signal is contained in e_{ijk}. The goal is to explain the highest variance in the original data by the PARAFAC2 components and thereby minimize e_{ijk}. Using an alternating least squares routine, the algorithm fits models until the improvement between iterations falls below a given convergence criterion. For this study the converge criterion was set to a relative change in errors by 10^{-9} or less.

Before PARAFAC2 was applied, larger retention time shifts between chromatograms were corrected so that the peaks of the internal standard CIN were aligned across all samples (peaks in interval X in Figure 1a). While this removes the majority misalignment due to retention time shifts between chromatograms (samples) there can still be minor remaining sample specific shifts. To circumnavigate issues associated with the considerable baseline drift due to changing solvent composition and elution of a broad DOM background signal, the chromatograms were transformed into their 2nd derivative (Figure 1b).

While in principle the PARAFAC2 analysis could be performed on the whole chromatogram across all samples, in practice this is very difficult, computationally slow, and often does not

provide robust results due to the very large number of components required. Rather, it is commonplace to divide the chromatogram into peak windows that are subjected to individual analysis (Amigo et al., 2010). In line with this approach, the chromatograms were divided into 10 intervals (Figure 1a&b) and each interval was characterized by PARAFAC2 independently following a routine created to achieve the best model solution (see Figure 2). For each interval, ten PARAFAC2 models were fitted with 1-4 components, which resulted in a total of 40 models per interval. The wavelength and the sample number mode were constrained to non-negativity, while the retention time mode was left unconstrained due to the nature of the 2nd derivative space. The stop criteria for the fitting of the models were set to 3000 iterations, and convergence required a relative change in fit of 10^{-9} or less. The PARAFAC2 modelling was performed in MATLAB (version R2021a) using the PLS toolbox (PLS_Toolbox 8.6.1, Eigenvector Research, Inc., Manson, WA).

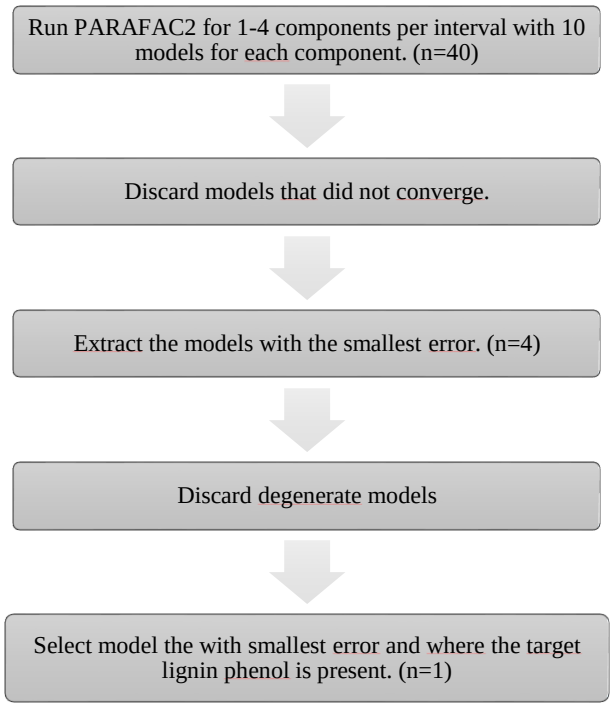


Figure 2: The PARAFAC2 routine is used to find the best model for all intervals. Degenerate models contain two components that are too similar spectrally (TCC>= 0.98) and the model may therefore suffer from overfitting.

To assess the spectral character of the components found in each model and compare them to the pure lignin phenol spectra, Tucker congruence coefficient (TCC) test was applied (Lorenzo-Seva & ten Berge, 2006; Tucker, 1951). A coefficient of 1 indicates that the spectra are perfectly similar, while a coefficient of 0 indicates that they are completely different. From the

40 models per interval, the best model was selected according to the scheme shown in Figure 2. First, all models that did not converge were discarded and secondly only the model with the highest variance explained, for 1 to 4 components respectively, was retained. From these remaining four models any model with two components that were similar because of overfitting and determined as TCC >= 0.98 were discarded since they are statistically invalid (Lorenzo-Seva & ten Berge, 2006; Rayens & Mitchell, 1997). The remaining model with the highest explained variance and a spectral loading with a TCC > 0.98 to the pure spectrum of the expected lignin phenol in that interval, was chosen as the final model. This process was subsequently automated to run as a script in MATLAB with little user intervention. The MATLAB script is available from Bruhn et al. (2021a).

Quantification of phenolic monomers

The PARAFAC2 sample scores were used (instead of peak area/height as done with manual integration) in the construction of calibration curves. The calibration curves ranged over three orders of magnitude (1 nM -25 µM). The range was therefore split into three segments, low (1, 2, 5, 10, 25 nM), middle (25, 50, 200, 1000 nM) and high (1, 2.5, 5, 10, 25 µM), and linear regression was performed for each of them. For calculation of lignin phenol concentrations in a specific sample, the appropriate segment was selected based on the PARAFAC2 scores. For the middle and high segments the calibration curves were forced through zero. The calibration standards ran in the middle of the sample run, one time.

The environmental concentration (nM) of each lignin phenol in the original seawater sample was calculated by multiplying the detected HLPC amount (nanomoles) by a concentration factor f_c:

$$f_c = \frac{V_{elute}}{V_{aliquot}} * \frac{V_{vial}}{V_{extract}}$$

Where V_{elute} is the volume eluted from the PPL cartridge (3 mL), $V_{aliquot}$ the volume of sample added to reaction vessel for cupric oxidation (0.6 mL), V_{vial} is the volume of the HPLC vial (0.200 mL) and $V_{extract}$ is the volume (mL) of original seawater passing through the PPL cartridge (see Table 1).

The concentration of the internal standard (CIN) measured in each sample was used to calculate the recovery of lignin phenols during the HLB extraction and clean-up. The recovery was calculated as the ratio between the detected concentration and the expected concentration of CIN (21.25 µM) in the HPLC vial. This recovery fraction was then used to correct for loss of natural lignin phenols in the samples by dividing the detected concentrations by it. We note that this entire calculation approach, from HPLC vial to environment, assumes no extraction bias originating from the SPE extraction using PPL (Arellano et al., 2018).

Total dissolved lignin phenol concentration (nM) was calculated as the sum of the 11 lignin phenol concentrations (TDLP11). The lignin carbon molar ratio (LCMR) was calculated as the ratio between lignin carbon and bulk DOC. For comparability with earlier studies, the TDLP11 was additionally normalized to the bulk DOC of the sample by dividing the TDLP11 in mg by the DOC in g (Λ_{11}: mg/g DOC).

Manual chromatogram integration

For comparison with the 2nd derivative/PARAFAC2 quantification method, two manual integration methods were performed on the chromatograms as well: peak height (apex method); and peak area (perpendicular drop method). Before performing the two manual integration methods, new baselines were created under the chromatographic peaks of the lignin phenols and the internal standard. The baselines were created by drawing a line between the two lowest points on each side of the peak. In case of co-eluting peaks, the baseline would be drawn between the two lowest points on each side of all peaks in the co-elution. The peak height in the apex method was measured as the distance between the newly created baseline and the apex of the peak (example in Figure S1). The perpendicular drop method instead draws lines to the new baseline at the point where the peak starts and ends (example in Figure S1). The area between these two lines is then calculated to determine the size of the peak. In case of co-elution with other peaks, the lowest point between the two peaks is used to draw the line. The peak height and peak area of each lignin phenol were determined using the extracted wavelength chromatogram at the wavelength of maximum absorbance for that specific lignin phenol (see Table 3 for the wavelength of maximum absorbance of all lignin phenols and the internal standard).

Both manual integration methods were applied to the raw analyte-specific wavelength chromatograms of the calibration standards, standard mixtures, the spiked sample and the same

sample without spike. The derived peak heights and peak areas were then used to plot calibration curves and from these calculate lignin phenol and internal standard concentrations following the same procedure as described for the 2nd derivative/PARAFAC2 method.

Limit of detection (LOD)

The LOD was calculated for each lignin phenol based on the low segment calibration curves (calibration standard no. 14-18, Table S1), by dividing the standard deviation of calibration curve residuals by the slope of the calibration curve and multiplying by three (Shrivastava & Gupta, 2011). For LOD determination in the presence of a DOM background, dilution of a natural seawater sample was performed by adjusting the injection volume in the HPLC instrument. For this a 1-100 fold dilution curve was used and here the dilution mimics the process of extracting less and less DOM from seawater onto the PPL cartridge. The LOD values were then calculated, similarly to the ones for the calibration curves, however based on the residuals of the dilution curve.

Assessment of quantification methods

The results from PARAFAC2 and the two manual integration methods were compared to each other by assessing sensitivity, specificity and recovery for each lignin phenol across the three quantification methods (PARAFAC2, apex method and perpendicular drop). Sensitivity comparison between the three quantification methods was assessed from the estimated LOD values for each. The specificity for the three quantification methods was assessed as the spectral similarity (TCC, see section "Selection of best PARAFAC2 solution") between the pure spectra and those extracted by the three different methods for each of the lignin phenols and the internal standard. For the perpendicular drop method, normalized spectra at each retention time point across the integrated lignin phenol peak were extracted and the TCC test was performed for each spectrum and a mean TCC value was then estimated for the whole peak. The recovery of the lignin phenols and the internal standard by the three quantification methods was examined as the ability to accurately find the added known concentration in a spiked sample (St. 11/19m mixed with the standard mixture). The spiked sample was repeatedly measured twelve times spread out across a run with 92 samples.

Comparative lignin phenol analysis on GC-MS

We compared results of our newly proposed method for analysis of lignin phenols using HPLC-DAD, with a more traditional method utilizing cupric oxide (CuO) oxidation, liquid-liquid extraction and derivatization, followed by gas chromatography-mass spectrometry (GC-MS), based on prior methods (Louchouarn et al., 2000; Benner and Kaiser, 2011 ; Osburn et al., 2016). For this comparison, we used water samples that were collected between 2019 and 2021 from surface waters of the Florida Coastal Everglades (FCE) and Florida Bay, FL. We hypothesized that if these divergent methods produced comparable lignin phenol concentrations and diagnostic ratios on coastal seawater samples, then the HPLC-DAD analysis is robust. Note that we made this comparison to test agreement between the methods and not to assess the accuracy of either method.

Statistical test and data availability

In order to test for differences between three quantification methods (PARAFAC2, apex method and perpendicular drop) and between groups of samples, analysis of variance (ANOVA) tests were performed with a significance level set to a p-value of 0.05. The ANOVA tests were performed in MATLAB (version R2021a) using the integrated anova1 function.

The MATLAB script for the 2nd derivative/PARAFAC2 algorithm is available from Bruhn et al. (2023a) and a dataset, including chromatograms from the oxidized DOM samples used in this study, can be downloaded from Bruhn et al. (2023b).

Assessment

Extracted DOC

The DOC concentration in the original seawater samples varied between 103-227 µM and 65-72 µM, respectively for Kattegat and Davis Strait (Table 1). Therefore, different sample volumes were used to extract DOM onto the PPL cartridges for the two locations, approximately 2.4 and 4.8 L from Kattegat and Davis Strait, respectively. The amount of DOC extracted onto the PPL cartridge was calculated from the difference in the DOC concentration between the originally sampled water and the permeate water (water after PPL extraction has been performed). The extracted amount varied between 50-312 µmol C for Kattegat and 107-245 µmol C for the Davis Strait (Table 2). Depending on the sample, either the whole extract

(for 7 samples) or a 20% aliquot (for 5 samples) was transferred to the reaction vessel for oxidation. This resulted in the amount of DOC in the reaction vials varying between 18-312 µmol C and 21-49 µmol C for Kattegat and the Davis Strait respectively (Table 2). The volume of the reaction vessel was 850 µL so the concentration of DOC during the oxidation varied between 22-367 mM for Kattegat and 25-57 mM for Davis Strait (Table 2).

Table 2: The calculated amount of DOC extracted by the PPL cartridge (based on measurement of DOC concentrations before and after extraction of the sampled seawater) and in the reaction vessel used for cupric oxidation. ND = no data.

Location		**PPL cartridge**	**In the reaction vessel**		
Station	**Depth (m)**	**DOC extracted (µmoles)**	**Aliquot**	**DOC (µmoles)**	**DOC (mM)**
3	100	107	0.2	21	25
4	100	203	0.2	41	48
5	100	245	0.2	49	58
6	5	ND	1	ND	ND
6	15	312	1	312	367
6	20	ND	1	ND	ND
11	4	ND	1	ND	ND
11	15	50	1	50	59
11	19	121	1	121	142
12	4	99	1	99	117
12	13	166	0.2	33	39
12	31	92	0.2	18	22

Chromatographic separation and PARAFAC2 decomposition

The 11 lignin phenols and the internal standard eluted in 10 intervals between 8.6 and 16.7 min (Figure 1a) in the order represented in Table 3. Five of the 12 compounds, more specifically PAD, VAD, SAD, PAL and CIN were successfully separated by the column. This is particularly clear in the standard mixture in ultrapure water (Figure 1a, black line). However, the phenols that eluted in intervals V-IX had variable degrees of co-elution. The spectral components of the final PARAFAC2 models for each of the intervals are shown in Figure 3. For nearly all intervals, more than one component was necessary to explain the elution profiles. For the intervals (VI-VIII) which overlapped, target phenol spectra were repeatedly isolated across the intervals. This can be seen for example by the reoccurrence of SAL and FAD in interval VI and VII, and by VAD and SAD in interval II and III (Figure 3). However, in contrast to interval V, where three target lignin phenols (CAD, PON and VAL) were isolated simultaneously, the models for SAL, FAD, VAD and SAD had a better fit when they were kept

separated in individual intervals. For nearly all intervals there were unidentified co-eluting compounds, which spectra did not resemble the character of a single analyte species.

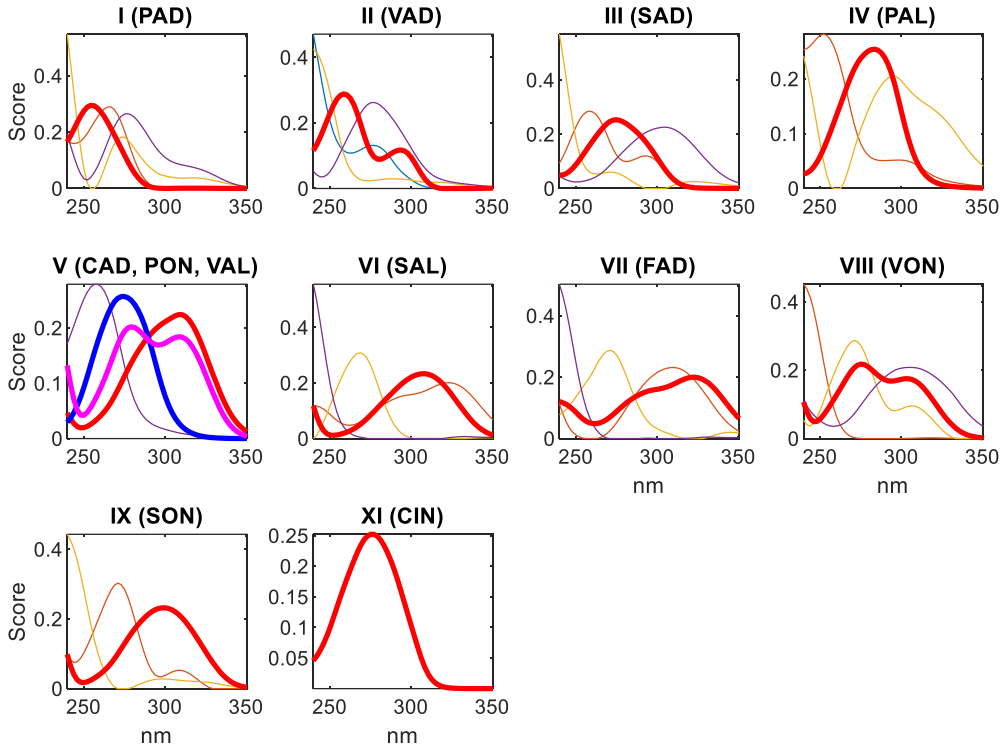


Figure 3: Spectral components found from the PARAFAC2 decomposition of each interval. The components that match the pure spectrum of the lignin phenols are indicated in bold red lines for each interval. For interval V, three phenols are included in the same PARAFAC2 model (blue=PON, red=CAD , magenta=VAL). Non-bold components either represent co-eluting phenols from neighboring intervals, unknown compounds or residual background (not removed by 2nd derivative transformation).

The PARAFAC2 representation of the chromatograms for each interval reflected the measured data very well for all the samples, reference solutions, and standards, with the final models explaining >98.7 % of the variance in the original data across all intervals. An example of the fit is shown in Figure 4, where the data in interval V is decomposed from three replicate measurements of the spiked sample. The actual model output is based on the 2nd derivative data, but the re-integrated data is also shown for a more intuitive representation (Figure 4, column to the right). The results demonstrate how the PARAFAC2 modelling succeed in splitting the co-eluting peaks into separate peaks in interval V (corresponding spectral components are shown in Figure 3) and excluded interference from the neighboring peaks. While the CAD elution profile follows a Gaussian-like shape (red profile in Figure 4d), PON and VAL (blue and magenta profiles in Figure 4d) indicate a degree of fronting. The fourth component

represents the background eluting signal and in this case has an absorption maximum at 258 nm (orange profile in Figure 4d). Using manual integration, this signal ends up being combined into the others.

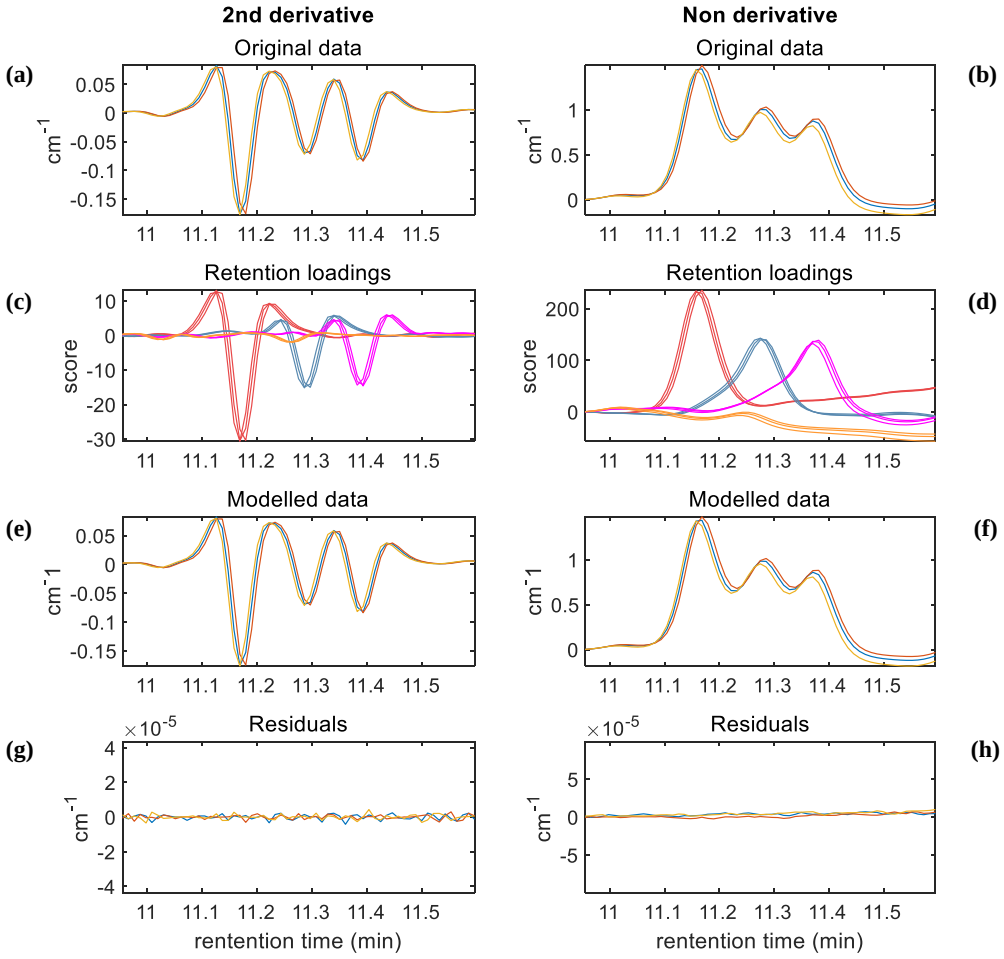


Figure 4: Best PARAFAC2 solution for data in interval V where the retention time profile is explained by four components, CAD (red), PON (blue), VAL (magenta) and unknown (orange), for three measurements of the spiked sample. Output from PARAFAC2 is provided in 2nd derivative form (c), however for ease of visualisation the output is also shown in its non-derivatized form (d). For both forms, the original data (a,b) is compared to modelled data (e,f) and the residuals (g,h) shows the data not explained by the model.

Molar absorptivity

Using the calibration standards, the molar absorptivity (ε_{max}) at the wavelength of maximum absorption (λ_{max}) of the phenolic monomers were calculated (Table 3). ε_{max} ranged between 3329-11067 L mol^{-1} cm^{-1} and λ_{max} ranged between 254 and 323 nm. It should be noted that ε_{max} is dependent on solvent, so the values reported here are specific for this mobile phase elution program. ε_{max} was calculated to give a first indication for which of the phenolic monomers

ought to have lowest LOD. However, this was not the case as CAD was found to have the second highest ε_{max} but far from the lowest LOD value (Table 4). Additionally, PAD had the lowest LOD value, but far from the highest ε_{max}. These contradictions are likely due to the additional factor of high co-elution with neighboring peaks.

Table 3: Overview of lignin phenols and spectral properties. The lignin phenols are presented in chromatographic order. The interval numerals refer to the highlighted sections in Figure 1. λ_{max} is the wavelength of maximum absorbance and ε_{max} is the molar absorptivity.

Interval	Lignin phenol name	Abbreviation	λ_{max}	ε_{max}
I	4-Hydroxybenzoic acid	PAD	254	9688
II	Vanillic acid	VAD	261	3900
III	Syringic acid	SAD	275	6900
IV	4-Hydroxybenzaldehyd	PAL	284	7550
V	p-Coumaric acid	CAD	309	11067
V	4-Hydroxyacetophenone	PON	275	8058
V	Vanillin	VAL	279	5092
VI	Syringaldehyde	SAL	309	4274
VII	Ferulic acid	FAD	323	11418
VIII	Acetovanillone	VON	275	5027
IX	Acetosyringone	SON	298	3329
X	Cinnamic acid	CIN	277	8134

Sensitivity, specificity and recovery

To assess the sensitivity of the three quantification methods (2^{nd} derivative/PARAFAC2 method, apex method and perpendicular drop method) the LOD for each of the lignin phenols and the internal standard were estimated. The instrumental LOD values for the 2^{nd} derivative/PARAFAC2 method were either comparable or better than the other two approaches for all lignin phenols (Table 4). The LOD values determined using calibration standards in ultrapure water however represent a best-case scenario. In contrast, natural water samples are prone to interference from the background matrix of DOM. The LOD values in the presence of a DOM background were on the whole higher than the pure water standards (Table 4), but PARAFAC2 still produced lower LOD values compared to the two manual integration methods, for most of the lignin phenols.

Table 4: Limit of detection for each lignin phenol estimated by the three different quantification methods: PARAFAC2; Apex; and Perpendicular drop. LOD values derived from pure water standards are estimated on from the regression of the calibration standards. Those derived from in the presence of background DOM are derived from a dilution series.

Limit of detection (nM)						
	Ultrapure water standards[a]			In presence of DOM background[b]		
	PARAFAC2	Apex	Perpendicular drop	PARAFAC2	Apex	Perpendicular drop
PAD	0.46	0.89	1.32	5.05	9.38	25.74
VAD	1.12	0.88	3.72	11.07	6.57	24.01
SAD	0.87	0.74	1.96	1.97	4.23	10.18
PAL	0.52	0.72	1.29	2.29	1.89	2.40
CAD	1.32	0.53	2.77	2.27	0.11	0.33
PON	2.36	3.16	0.13	3.82	11.10	4.81
VAL	0.84	0.01	0.75	1.05	16.73	22.68
SAL	2.77	2.67	4.79	1.40	5.35	10.30
FAD	1.58	0.64	0.96	0.50	0.57	1.51
VON	1.96	0.62	1.64	3.49	5.05	11.19
SON	2.32	3.87	5.16	0.90	5.19	7.74
CIN	1.54	0.44	0.86	4.11	10.13	8.84

To assess the specificity of the three quantification methods, the spectral similarity, between the extracted spectra from each of the methods was assessed (Table 5, Figure S2). While the spectra derived from the PARAFAC2 models matched the pure standards (due to the procedure in Figure 2), the spectra from the manual integration approaches often deviated, in particular if the sample was not spiked (Table 5).

Table 5: Spectral similarity (TCC, where 1 means completely similar) between the pure spectra and the ones extracted from the three different integration methods. For the perpendicular drop method all spectra at each retention point (n=number of retention points) across the integrated lignin phenol peak are extracted, normalized and the TCC test is then performed for each of the spectra to that of the pure lignin phenol spectrum, where a mean TCC is calculated in the end.

Tucker congruence coefficient					
	Across all samples	Spiked sample		Natural sample	
	PARAFAC2	Apex	Perpendicular	Apex	Perpendicular
PAD	1.00	1.00	1.00 ± 0.00 (n=12)	1.00	1.00 ± 0.00 (n=07)
VAD	0.99	1.00	0.92 ± 0.08 (n=19)	0.73	0.84 ± 0.09 (n=19)
SAD	1.00	1.00	0.89 ± 0.16 (n=25)	0.90	0.73 ± 0.22 (n=25)
PAL	1.00	1.00	0.85 ± 0.19 (n=33)	0.55	0.59 ± 0.07 (n=17)
CAD	1.00	1.00	0.95 ± 0.07 (n=13)	0.53	0.44 ± 0.07 (n=12)
PON	1.00	1.00	0.98 ± 0.02 (n=09)	0.84	0.84 ± 0.05 (n=09)
VAL	1.00	1.00	1.00 ± 0.00 (n=09)	1.00	0.99 ± 0.01 (n=09)
SAL	1.00	0.99	0.99 ± 0.01 (n=15)	0.56	0.32 ± 0.07 (n=15)
FAD	1.00	1.00	1.00 ± 0.00 (n=15)	0.78	0.77 ± 0.10 (n=07)
VON	1.00	0.98	0.96 ± 0.03 (n=18)	0.45	0.44 ± 0.03 (n=13)
SON	1.00	1.00	0.98 ± 0.03 (n=20)	0.79	0.77 ± 0.10 (n=20)
CIN	1.00	1.00	0.96 ± 0.11 (n=35)	1.00	0.95 ± 0.09 (n=35)

This questions the validity of using the manual integration approach, without further method optimization for better peak separation. Despite the apparent relatively good performance of the LODs for the manual approaches, they are clearly not isolating the signal from the specific phenols (Figure S2).

The recovery of lignin phenol concentration for spiked samples using the 2nd derivate/PARAFAC2 method, varied between 91-101 % (Table 6).

Loss of lignin phenols during HLB extraction and clean-up

The loss of the lignin phenols and the internal standard purely due to the performance of the HLB cartridge was tested by passing three standard mixture samples (ultrapure water and no cupric oxidation) through the HLB cartridge. The mean recovery on the HLB cartridge was 82% (± 19%) across all lignin phenols and the internal standard (Table 6). As CIN showed to have a mean recovery of 86% (± 19%), using CIN as an internal standard for the HLB extraction and subsequent clean-up steps corrects very well for the loss of the lignin phenols. The effect of the purification using the HLB cartridge was monitored across 44 natural samples (including samples from an unpublished dataset containing marine samples) and the average losses of CIN for these samples was found to be 86% (± 11%) as well. Of the 44 samples, only four samples exhibited losses higher than 40% for CIN. On the whole, the recovery of lignin phenols during the purification with HLB and the recovery of CIN in environmental samples were found to be very similar to the findings from other studies (Arellano et al., 2018; Kaiser & Benner, 2012)

Table 6: The recovery of known concentrations of added lignin phenols in a natural sample (St. 11/19m). The spiked sample was injected 12 times during the whole sample batch run and quantified using the PARAFAC2 approach. The second column shows the recovery of each lignin phenol during HLB clean up, calculated as the difference between the measured concentration and the expected concentration (20 µM) in three replicate standard mixtures (n=3).

	PARAFAC2	**HLB clean-up**
Compound	**Recovery (%)**	**Recovery (%)**
PAD	99.8 ± 2.08	89.6 ± 15.8
VAD	91.0 ± 1.94	80.6 ± 16.2
SAD	95.9 ± 1.22	89.1 ± 18.0
PAL	93.8 ± 1.41	ND
CAD	98.0 ± 1.35	84.8 ± 14.6
PON	97.9 ± 1.16	86.0 ± 18.6
VAL	101.1 ± 0.63	67.5 ± 24.3
SAL	93.7 ± 2.91	83.1 ± 19.4
FAD	99.6 ± 1.64	89.1 ± 16.7
VON	98.8 ± 1.24	83.0 ± 23.4
SON	97.6 ± 1.45	63.4 ± 14.5
CIN	93.5 ± 1.47	85.9 ± 18.8

Phenol concentration and composition in samples

The lignin phenol concentrations and indices for the seawater samples are shown in Table 7. For Kattegat, the TDLP11 ranged between 4.3 - 10.09 nM, while in the Davis Strait TDLP11 was lower and ranged between 2.1 - 3.2 nM. Typical for all samples was that the concentration of PAD, VAD and VAL was highest ranging from 1.12 – 2.54 nM, 1.00 - 2.53 nM, and 0.69 - 1.63 nM, respectively in Kattegat, and 0.53 - 0.57, 0.50 - 0.77, and 0.14 - 0.59, respectively in the Davis Strait samples. The ratio between S (SAD, SAL, SON) and V (VAD, VAL, VON) phenols (S/V) for Kattegat and Davis Strait ranged between 0.15-0.26 and 0.19-34, respectively, while the ratio between C (CAD, FAD) and V phenols (C/V) were comparable between the two sites ranging between 0.08 - 0.17. The acid (Ad) to aldehyde (Al) ratios for S phenols (Ad/Al (S)) for Kattegat ranged between 1.52- 2.52, which generally was higher to that detected in the Davis Strait samples (1.33 - 1.64), while the equivalent ratio for V phenols (Ad/Al (V)) differed slight between the two sites 0.92 – 1.98 in Kattegat and 1.29 - 3.73 in Davis Strait). Carbon normalized TDLP11 values (Λ_{11}) were similar between Kattegat and Davis Strait, ranging between 0.4-0.9 mg/g DOC. The similarity in Λ_{11} implies that the fractions of terrestrial DOC are similar in both water masses. The LCMR, ratio between lignin carbon and bulk DOC, ranged from 0.03 - 0.07 $\times 10^{-3}$.

To assess the reproducibility from cupric oxidation to analysis, we examined three replicate samples from the SRNOM stock solution (concentrations and diagenetic values are shown in Table 8). The mean TDLP11 for the three replicate SRNOM samples was found to be 293.04 μmol mol^{-1} C with a relative standard deviation (RSD) between replicates of 2.46 %. Most of the lignin phenols had concentrations ranging between 2.43-26.24 μmol mol^{-1} C and RSD between 2.03-16.84%, while only PAD and VAD reached as high as 72.85 (± 11.22%) and 100.35 (± 2.70%) μmol mol^{-1} C. The S/V, C/V, Ad/Al (S) and Ad/Al (V) ratio was 0.24 (± 1.83%), 0.14 (± 3.24%), 6.06 (± 3.24%) and 9.65 (± 13.68 %) respectively. Both the S/V and C/V for SRNOM was comparable to the seawater samples, while both the Ad/Al (V) and Ad/Al (S) were notably higher. The SRNOM reference showed to yield a Λ_{11} of 3.81 mg/g (± 2.46%), which is 4-10 times higher than the seawater samples reflecting the higher fraction of lignin compared to oceanic DOM, as expected.

Table 7. Concentrations of individual lignin phenols for seawater samples (nM) and calculated diagenetic ratios.

Station	St. 3	St. 4	St. 5	St. 6			St. 11			St. 12		
Depth	**100 m**	**100 m**	**100 m**	**5 m**	**15 m**	**20 m**	**4 m**	**15 m**	**19 m**	**4 m**	**13 m**	**31 m**
PAD	0.57	0.59	0.53	2.54	2.36	1.80	1.58	1.67	1.12	3.53	2.31	1.24
VAD	0.77	0.50	0.50	1.35	1.62	1.58	0.88	1.40	1.00	2.19	2.53	1.74
SAD	0.12	0.13	0.10	0.26	0.34	0.30	0.24	0.31	0.23	0.45	0.39	0.24
PAL	0.33	0.25	0.24	0.64	0.64	0.64	0.47	0.61	0.48	0.80	0.94	0.74
CAD	0.11	0.06	0.05	0.33	0.17	0.17	0.27	0.13	0.10	0.23	0.26	0.23
PON	0.29	0.17	0.19	0.60	0.57	0.54	0.68	0.45	0.30	0.86	0.77	0.60
VAL	0.59	0.14	0.29	0.98	1.01	1.27	0.96	0.69	0.74	1.11	1.54	1.63
SAL	0.09	0.08	0.06	0.17	0.14	0.13	0.13	0.12	0.11	0.18	0.18	0.16
FAD	0.04	0.03	0.03	0.08	0.10	0.09	0.06	0.08	0.06	0.13	0.12	0.08
VON	0.11	0.10	0.23	0.04	0.36	0.34	0.01	0.17	0.10	0.49	0.46	0.26
SON	0.18	0.04	0.04	0.14	0.10	0.10	0.10	0.11	0.10	0.12	0.13	0.15
S [a]	0.38	0.25	0.20	0.57	0.58	0.53	0.48	0.54	0.43	0.76	0.70	0.55
V [b]	1.47	0.73	1.03	2.37	2.99	3.20	1.85	2.25	1.84	3.79	4.53	3.63
C [c]	0.15	0.09	0.09	0.41	0.27	0.26	0.33	0.21	0.16	0.37	0.37	0.31
TDLP11 [d]	3.19	2.08	2.26	7.13	7.41	6.96	5.38	5.73	4.34	10.09	9.62	7.08
LMCR [e]	0.05	0.03	0.03	ND	0.03	0.04	ND	0.04	0.04	0.06	0.06	0.07
Λ_{11} [f]	0.64	0.39	0.40	ND	0.41	0.56	ND	0.49	0.49	0.71	0.75	0.88
S/V	0.26	0.34	0.19	0.24	0.19	0.17	0.26	0.24	0.23	0.20	0.15	0.15
C/V	0.10	0.13	0.08	0.17	0.09	0.08	0.18	0.10	0.09	0.10	0.08	0.08
Ac/Ad (S) [g]	1.33	1.64	1.58	1.52	2.38	2.26	1.81	2.52	2.11	2.50	2.23	1.51
Ac/Ad (V) [g]	1.29	3.73	1.71	1.38	1.61	1.24	0.92	2.04	1.34	1.98	1.64	1.07

[a] Sum of syringyl phenols (SAD, SAL, SON)

[b] Sum of vanilyl phenols (VAD, VAL, VON)

[c] Sum of cannimyl phenols (CAD and FAD)

[d] Total dissolved lignin phenols

[e] Lignin carbon molar ratio (10^{-3})

[f] Carbon normalized (mg/g DOC)

[g] Ac=acids, Ad=aldehydes

ND = No data

Table 8: Lignin phenol concentrations and calculated diagenetic ratio for Suwannee River natural organic matter (SRNOM).

Compound	SRNOM (umol/mol carbon) (n=3)	RSD (%)
PAD	72.85 ± 8.18	11.22
VAD	100.35 ± 2.70	2.70
SAD	14.73 ± 0.22	1.47
PAL	26.24 ± 0.75	2.87
CAD	12.69 ± 0.29	2.27
PON	25.60 ± 1.48	5.77
VAL	10.51 ± 1.14	10.83
SAL	2.43 ± 0.05	2.03
FAD	4.20 ± 0.08	1.96
VON	10.93 ± 1.84	16.84
SON	12.50 ± 0.34	2.73
S [a]	29.67 ± 0.21	0.67
V [b]	121.79 ± 1.49	1.22
C [c]	16.89 ± 0.33	1.98
TDLP11 [d]	293.04 ± 7.28	2.46
LMCR [e]	0.29 ± 0.01	2.46
Λ_{11} [f]	3.81 ± 0.08	2.46
S/V	0.24 ± 0.00	1.83
C/V	0.14 ± 0.00	3.18
Ac/Ad (S) [g]	6.06 ± 0.20	3.24
Ac/Ad (V) [g]	9.65 ± 1.32	13.68

[a] Sum of syringyl phenols (SAD, SAL, SON)

[b] Sum of vanilyl phenols (VAD, VAL, VON)

[c] Sum of cannimyl phenols (CAD and FAD)

[d] Total dissolved lignin phenols

[e] Lignin carbon molar ratio (10^{-3})

[f] Carbon normalized (mg/g DOC)

[g] Ac=acids, Ad=aldehydes

Assessment of effect of DOC concentration on lignin phenols and ratios

To investigate if there was an effect of DOC concentration in the reaction vessel on the environmental lignin phenol concentrations and diagenetic ratios, all the parameters were plotted against the DOC concentration for the Kattegat samples (Davis Strait samples were excluded to avoid a location bias). From Figure 5 & 6, no visual correlation with increasing DOC concentration in the reaction vessel was observed for any of the parameters. Instead, the samples from Kattegat were divided into two groups, samples with less (n=3) and samples with

more (n=3) than the mean DOC concentration (60 mM) in the reaction vessel (see the black line in Figure 5 & 6 for division of samples) to perform ANOVA between the two groups.

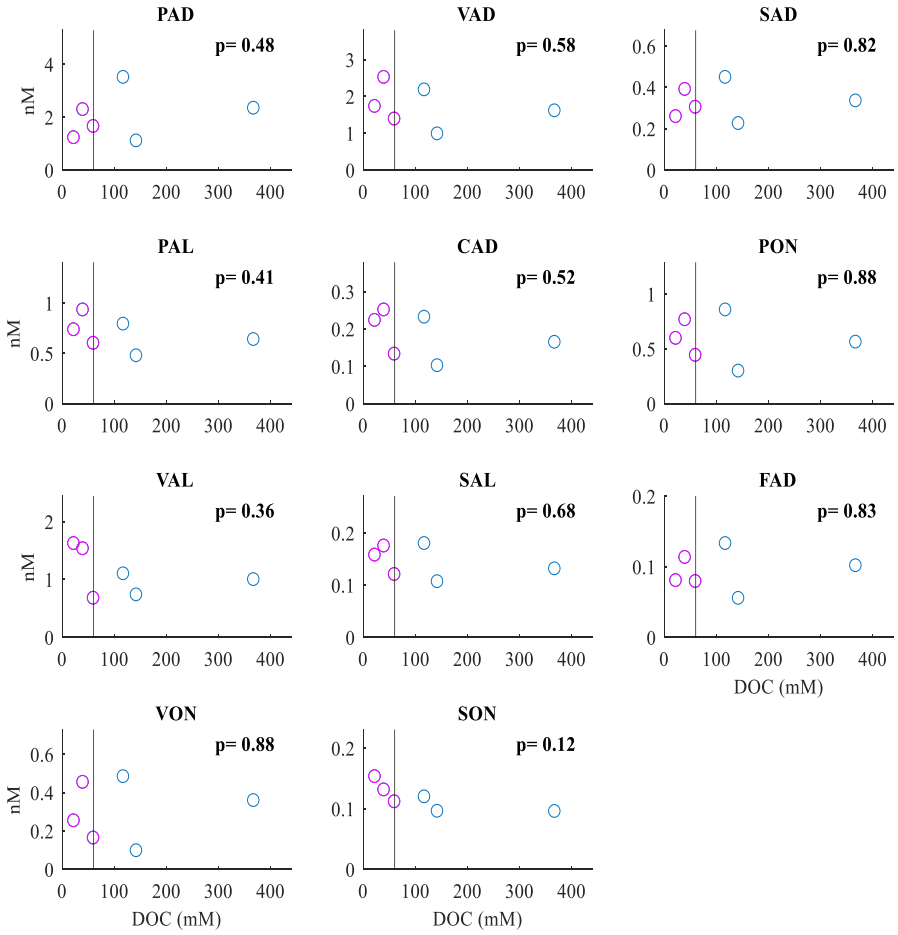


Figure 5: Lignin phenol concentrations (nM) in seawater plotted against the DOC concentration (mM) in the reaction vessel for samples collected in Kattegat. The black line indicates the mean DOC concentration used to divide the samples into two groups (purple and blue) used for statistical analysis (ANOVA one-way test).

Comparing lignin phenols at their environmental concentration (nM) showed no statistical difference between the two groups as an effect of DOC concentration (p-values between 0.12-0.88; see Figure 5). Similarly, TDLP11 did not differ significantly between the two groups (p-value=0.92; Figure 6) and neither did S/V, C/V, Ad/Al (V) and Ad/Al (S) (p-values between 0.44-0.87; see Figure 6).

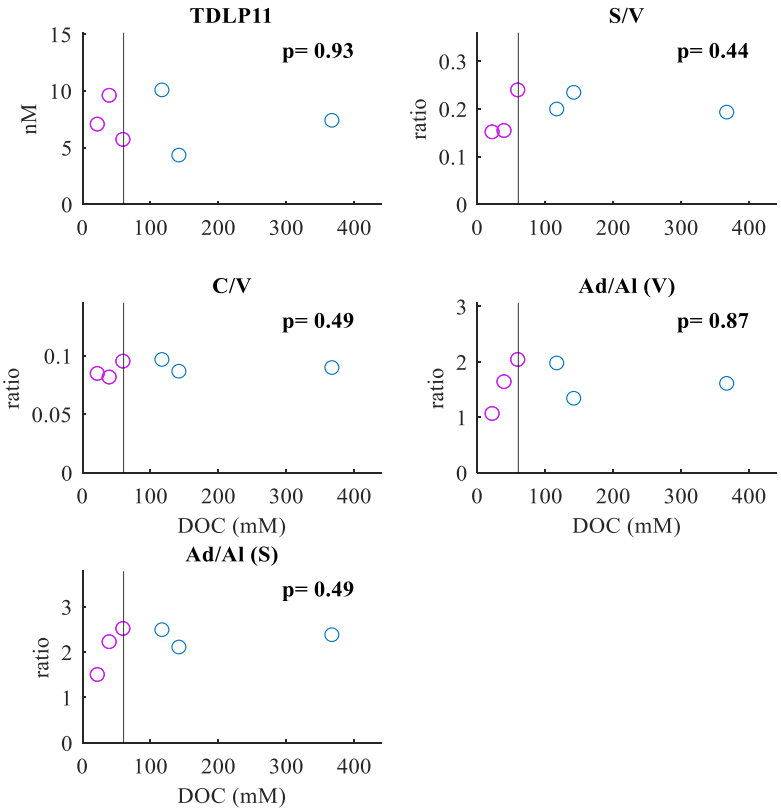


Figure 6: TDLP11 (nM) and diagenetic ratios (unitless) plotted against the DOC concentration (mM) in the reaction vessel for samples collected in Kattegat. The black line indicates the mean DOC concentration used to divide the samples into two groups (purple and blue) used for statistical analysis (ANOVA one-way test).

Comparison to lignin phenols obtained from GC-MS

While total lignin phenol concentrations obtained by the machine-learning assisted HPLC-DAD method were found to be similar to that determined with an established GC-MS method, there were clear differences in the composition (Figure 7). SAL, SON and SAD concentrations compared well between the two oxidation and quantification methods, but there were clear differences in the vanillyl and cinnamyl phenols. The acid, aldehyde and ketone composition of the vanillyl phenols differed, however the sum of vanillyl phenols compared well.

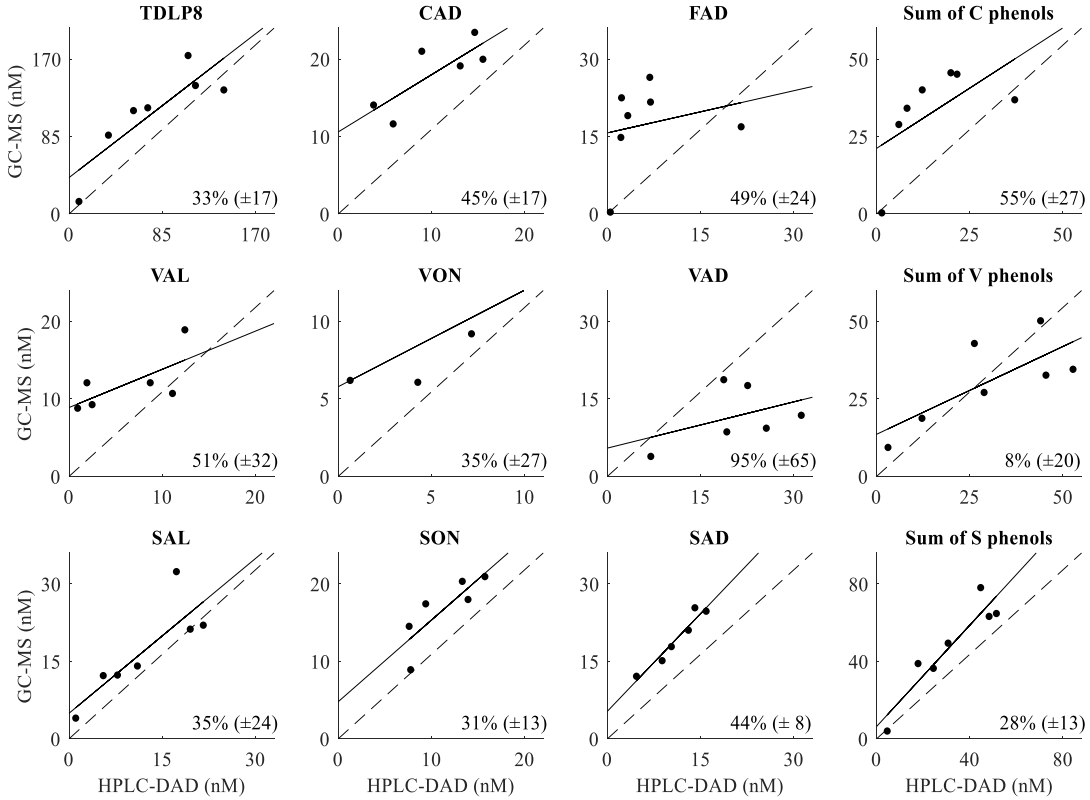


Figure 7: Comparison between quantified lignin phenol concentrations, for FCE samples, obtained by two completely different methods, where the HPLC-DAD corresponds to the newly proposed machine-learning assisted method in this study. A complete description of both methods can be found in the method section. Each of the dots corresponds to one FCE sample, the dashed line indicate a 1:1 relationship between the two methods, and the solid line indicates the linear regression between the dots. The percentage in each window indicate how much the concentration across the seven FCE samples obtained with the HPLC-DAD method on average differed from the ones obtained by GC-MS method, with standard deviation in parentheses.

Discussion

The majority of the lignin phenols, including the internal standard, were fully resolved in the standard mixtures due to the HPLC separation, with exception of the three lignin phenols (CAD, PON, VAL) in interval V. The 2nd derivative transformation and PARAFAC2 decomposition helped improve the separation without the necessity of increasing the complexity of the chromatographic method. In contrast, Fischer & Höffler (2021) included multiple elution steps to achieve proper separation of all lignin phenols. Our HPLC separation reduces the run time by 20 min compared to the HPLC-DAD method developed by Lobbes et al. (1999) and is overall similar in run time to the more recent HPLC-DAD method developed by Fischer & Höffler (2021).

There are considerable benefits with respect to sensitivity and spectral confirmation of analytes using PARAFAC2. Skov & Bro (2008) showed that applying PARAFAC2 in contrast to the automated manual integration algorithm provided by ChemStation (software from Agilent Scientific) increased the linearity for co-eluting peaks and thereby potentially improved LOD. From the LOD values in Table 4 it was apparent that, for most of the lignin phenols in ultrapure water the sensitivity was higher using the 2nd derivative/PARAFAC2 method compared to the two manual integration methods. The higher sensitivity for ultrapure water standards is likely due to the fact that the method removes the influence of instrumental noise which is important at low signal intensities (Amigo et al., 2008, 2010).

The presence of a DOM background overall resulted in a lower sensitivity for all three quantification methods, based on the estimated LOD values. While extrapolating baselines in the manual integration methods is an alternative approach to circumnavigate the signal from background DOM remaining in the sample after the HBL clean-up, the 2nd derivative/PARAFAC2 method is a far simpler procedure requiring little effort from the operator. In addition, the PARAFAC2 decomposition can better separate interferences of potential co-eluting unknown analytes from the final estimation of lignin phenol concentrations, and therefore lead to increased sensitivity (i.e. lower LOD, Table 4). Achieving lower LOD values in the presence of DOM, with the 2nd derivative/PARAFAC2, also implies that the volume of original seawater sample required for samples for seawater can be reduced compared to the use of manual integration techniques. Overall, our observations between manual integration and PARAFAC2 modelling of chromatographic data underline the findings of Amigo et al. (2008, 2010) where they also showed that resolving highly overlapping peaks in chromatographic data by PARAFAC2 is a very powerful tool. Additionally, there is a considerable improvement in data processing, removing the need for manual baseline calculation and integration of each peak, and potential for user bias.

From Table 5 and Figure S2 it is clear that the specificity of the 2nd derivative/PARAFAC2 method exceeds the manual methods. The higher specificity for identification of lignin phenols using the 2nd derivative/PARAFAC2 probably comes from the mathematical separation of co-eluting analytes and removal of background noise, which for manual integration cannot be removed properly and therefore interferes with the extracted spectrum of the lignin phenol, as seen in Figure S2 (red lines).

Comparison to existing methods

The calculated LOD values for the 2nd derivate/PARAFAC2 method were an order of magnitude lower, for most of the lignin phenols, compared to previous studies with a similar analytical setup (Fischer & Höffler, 2021; Lobbes et al., 2000; Steinberg et al., 1984). The LOD values achieved are comparable to newer MS studies on lignin quantification where Reuter et al. (2017) found LOD values between 1-3 nM and Yan & Kaiser (2018b) found between 1-8.7 nM (10-87 femtamoles in 10 µL injection volume). The RSD for the triplicate SRNOM samples for the whole process, from cupric oxidation to HPLC analysis, was found to range between 2.03-16.84%, with most of the lignin phenols having a RSD below 3%. Yan & Kaiser (2018b) found RSD for their Suwannee River humic acid standard to be ranging between 1.1-14.9% for the lignin phenols produced from their whole process, with most of the phenols having a RSD over 5%, using an ultra-HPLC-MS/MS setup. From a sensitivity and reproducibility aspect, the proposed HPLC-DAD combined with 2nd derivative/PARAFAC data decomposition is therefore competitive with more advanced instrumentation. One of the disadvantages of absorbance compared to MS detection has always been lower specificity. However, application of the 2nd derivative/PARAFAC2 method and spectral matching of isolated components clearly improves specificity of HPLC-DAD.

Effects of organic matter content on oxidation

Yan & Kaiser (2018a) found that using DOM containing less than 100 µg DOC (8.33 µmoles) in their reaction vessel (220 µL; 38 mM DOC) did not only result in substantial lower TDLP values, but also changed the composition and thereby the calculated ratios between lignin phenols, due to over-oxidation. However, they found that adding ascorbic acid and lowering the concentration of NaOH, provided more stable results, for DOC amounts down to 5 µg (0.42 µmoles; 2 mM DOC in the reaction vessel). From their study, it however seems that there may also be an upper limit, with changes in lignin phenol composition at DOC concentrations higher than 38 mM, in the reaction vessel. Kaiser & Benner (2012) also found, using a slightly different oxidation method, that a minimum of 50 mM DOC was required in the reaction vessel to avoid over-oxidation problems and that an increase to 133 mM DOC did not change lignin phenol composition nor TDLP values, however recommend the addition of glucose as antioxidant for samples with less than 166 mM DOC. From our study, the statistical test between samples with less and more than 60 mM DOC in the reaction vessel, indicated no significant difference in lignin phenol concentrations and ratios between the two groups (Figure

5 and 6). These findings suggest no systematic effect on lignin phenol yields of high DOC concentration in the reaction vessel and that the $CuSO_4$ concentrations in the reaction vessel with the current method, similar to that proposed by Yan & Kaiser (2018a), is enough to oxidize large amounts of DOM (up to 327 mM). This was further confirmed with a larger dataset of field samples (see Supplementary Information, Figure S3 & S4). These findings indicate that concerns with DOC concentration in the reaction vessel should mostly be focused on avoiding too small sample size, which however appear to be countered with the addition of an antioxidant.

Comparison with measurements from other methods.

Overall, the findings in our study agrees with other studies, that the sum of lignin phenols makes < 1 % of DOC in coastal waters (Harvey & Mannino, 2001; Hernes & Benner, 2003; Osburn & Stedmon, 2011; Walker et al., 2009). The concentration of lignin phenols found in this study for Kattegat are lower than what has previously been published for the same area. Osburn & Stedmon (2011) found that the sum of S and V phenols (S+V) in Kattegat ranged from 0.97 - 5.08 µg/L, where we found the S+V to range between 0.13 – 0.29 µg/L. The S/V and Ad/Al (V) ratios for Kattegat agree with that found by Osburn & Stedmon (2011). Similar to Osburn & Stedmon (2011) we also found an inverse relationship between lignin and salinity (see Figure S5). For the Davis Strait, the individual lignin phenol concentrations varied between 0.04 and 0.77 nM, while TDLP11 varied between 2.08-3.19 nM, and fit well with other studies in similar regions. Kaiser & Benner (2012) reported 1.9 nM for TDLP11 in Arctic Ocean surface waters and Benner et al. (2005) found approximately 2 nM for S + V for the Arctic outflow on East Greenland. The S/V ratios (0.19-0.34) are similar to that reported by Benner et al. (2005) for the East Greenland current (0.28) and by Kaiser & Benner (2012) for Arctic Ocean surface waters (0.32).

The comparison of lignin phenol concentrations obtained with those from a well-established GC-MS method that utilizes a different oxidation and clean-up procedure was promising. Overall, both methods showed similar results, in particular for the concentration of TDLP8, individual S phenols, sum of S phenols and sum of V phenols. Interlaboratory comparison between HPLC-DAD and GC-MS was also performed by Lobbes et al. (1999), where they found substantial differences for most of the lignin phenols (as much as 300% difference for PAD). Their outcomes were however similar to the findings here in our study revealing smaller

differences when comparing the sum of lignin phenols between HPLC-DAD and GC-MS, instead of individual lignin phenols. Yan & Kaiser (2018b) showed when comparing HPLC and GC coupled with tandem MS, the deviations in oxidation and chromatography only led to differences of 0-16%.

That we were able prove similarities between the HPLC-DAD and GC/MS methods in quantifying lignin phenols holds promise for the future of our new method, though clearly future work is required to identify reasons for these uncertainties. We speculate that the difference in concentrations obtained by the HPLC-DAD and GC-MS methods are mostly due to the oxidation (respectively using $CuSO_4$ and CuO) and preparation steps rather than the detection. Yan & Kaiser (2018a) discovered that the oxidation of DOM with $CuSO_4$ compared to CuO mostly led to similar or higher concentration of lignin phenols, and they found that this was clearest for V phenols. In our study, we only observed a higher concentration of VAD measured by the HPLC-DAD method using $CuSO_4$, compared to the GC-MS method using CuO, whereas we rather observed similar or underestimation for the rest of the lignin phenol concentrations. This possibly indicates an over oxidation occurring and clearly warrants further investigation and interlaboratory comparison.

Future perspectives.

The seawater sample volume requirements for HPLC-DAD have been greatly reduced compared to earlier lignin quantifications due to the improved LOD values accomplished by the 2nd derivative/PARAFAC2 method and additionally with the use of small reaction volumes for the cupric oxidation. Although as much as 5 L for certain samples was extracted, the resulting lignin phenol concentrations indicated that 1 L would have certainly provided sufficient phenol yields for quantifications. The extraction volume needed will however depend on sample area, as the composition and concentration of DOM and hence lignin phenols can vary greatly. It is therefore necessary to investigate what the minimum volume needed for extraction is in a particular ocean region. The need of smaller sample volumes will lower the water budget on research expedition and cut the run time for SPE extraction, reducing cost and equipment needed. The lower sample volume also means that lignin measurements can be carried out for small-scale experiment designs, i.e. bio- and photodegradation experiments. Additionally switching to an ultra-HPLC setup can lower solvent usage even more and at the

same decrease run time by increasing pressure, while maintaining resolution (Yan & Kaiser, 2018b).

A challenge with PARAFAC2 is often the required skills of the user to choose the right amount of components. However, with the routine proposed here (see Figure 2) the choice is now automated and this should allow for wider use in lignin quantification using HPLC-DAD chromatography. Also the application of PARAFAC2 here considerably reduces the time needed to process the chromatographic data, and limits user bias. This is not the first demonstration of automation of PARAFAC2 in chromatography as Johnsen et al. (2014) has also shown that they could automate the selection of the right model using a classification model (partial least squares-discriminant analysis, PLS-DA) based on seven quality criteria. However, a recent integrated approach titled PARAFAC2 based Deconvolution and Identification System (PARADISe) has compromised the complex coding and thereby made the application of PARAFAC2 extremely user-friendly, timesaving, and showed to produce reliable results that are less user-dependent (Petersen & Bro, 2018). Additionally, Baccolo et al., (2021) has automated the application of PARADISe for untargeted GC-MS analysis, so that it can be applied over the entirety of the chromatogram and extract all relevant spectra, elution profiles and relative abundance of different components, which can be attributed to specific compounds.

Comments & recommendations

Application of the 2nd derivative/PARAFAC2 method to HPLC-DAD chromatograms overall showed to increase performance at lower concentrations, leading to a lowering of LOD values. The 2nd derivate/PARAFAC2 method also improved specificity and thereby confidence in identification of targeted lignin phenols for measurements of natural samples, with a DOM matrix background. The approach opens the opportunity to perform lignin quantification using HPLC-DAD with little effort required for optimising chromatographic conditions and integrating chromatograms, since it facilitates the separation of highly overlapping peaks that are otherwise challenging to isolate. Applying the proposed PARAFAC2 algorithm to 2nd derivative chromatograms therefore speeds up the process and reduces interference from background DOM matrix. The interlaboratory comparison further showed that HPLC-DAD with our 2nd derivative/PARAFAC2 method can achieve similar results to GC-MS. This reinforces the use of both techniques for the quantification of lignin phenols.

It was found that adding varying sample size, hence concentrations of DOC, to the reaction vessel did not significantly change the lignin phenol concentration nor diagenetic ratios across samples from Kattegat. This suggests that our cupric oxidation method can be applied to a variety of DOC concentrations without much concern. However, the DOC limits should still be tested for the specific sample area and laboratory conditions.

Our 2nd derivative/PARAFAC2 method provides scientists and laboratories with an option to measure lignin in the ocean, even within a complex DOM matrix, using a simple HPLC-DAD setup instead of mass spectrometry. This alternative is cost-effective and eliminates the need for derivatization of the DOM prior to analysis. Finally, the 2nd derivative/PARAFAC2 method provides a powerful way to enhance analytical measurements that span aquatic environments and scientific disciplines holding promise for other applications that provide similar data structure such as the quantification of amino acids, and algal pigments.

References

Amigo, J. M., Popielarz, M. J., Callejón, R. M., Morales, M. L., Troncoso, A. M., Petersen, M. A., & Toldam-Andersen, T. B. (2010). Comprehensive analysis of chromatographic data by using PARAFAC2 and principal components analysis. Journal of Chromatography A, 1217(26), 4422–4429. https://doi.org/10.1016/J.CHROMA.2010.04.042

Amon, R. M. W., Budéus, G., & Meon, B. (2003). Dissolved organic carbon distribution and origin in the Nordic Seas: Exchanges with the Arctic Ocean and the North Atlantic. Journal of Geophysical Research: Oceans, 108(C7), 3221. https://doi.org/10.1029/2002JC001594

Arellano, A. R., Bianchi, T. S., Hutchings, J. A., Shields, M. R., & Cui, X. (2018). Differential effects of solid-phase extraction resins on the measurement of dissolved lignin-phenols and organic matter composition in natural waters. Limnology and Oceanography: Methods, 16(1), 22–34. https://doi.org/10.1002/lom3.10224

Baccolo, G., Quintanilla-Casas, B., Vichi, S., Augustijn, D., & Bro, R. (2021). From untargeted chemical profiling to peak tables – A fully automated AI driven approach to untargeted GC-MS. TrAC Trends in Analytical Chemistry, 145, 116451. https://doi.org/10.1016/J.TRAC.2021.116451

Benner, R., & Kaiser, K. (2011). Biological and photochemical transformations of amino acids and lignin phenols in riverine dissolved organic matter. Biogeochemistry, 102(1), 209–222. https://doi.org/10.1007/S10533-010-9435-4/FIGURES/7

Benner, R., Louchouarn, P., & Amon, R. M. W. (2005). Terrigenous dissolved organic matter in the Arctic Ocean and its transport to surface and deep waters of the North Atlantic. Global Biogeochemical Cycles, 19(2), 1–11. https://doi.org/10.1029/2004GB002398

Bro, R. (1997). PARAFAC. Tutorial and applications. Chemometrics and Intelligent Laboratory Systems, 38(2), 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4.

Bruhn, A. D., Wünsch, U., & Stedmon, C. A. (2023a). 2nd derivative/PARAFAC2 algorithm for lignin phenol analysis from HPLC-DAD chromatograms of oxidized ocean DOM. DTU Data (data.dtu.dk). doi: 10.11583/DTU.22178825.

Bruhn, A. D., Wünsch, U., & Stedmon, C. A. (2023b). HPLC-DAD chromatrograms of oxidized oceanic DOM used for lignin phenol analysis. DTU Dsata (data.dtu.dk). doi: 10.11583/DTU.22182694.

Ciais, P., Borges, A. v., Abril, G., Meybeck, M., Folberth, G., Hauglustaine, D., & Janssens, I. A. (2008). The impact of lateral carbon fluxes on the European carbon balance. Biogeosciences, 5(5), 1259–1271. https://doi.org/10.5194/BG-5-1259-2008

Cole, J. J., Prairie, Y. T., Caraco, N. F., McDowell, W. H., Tranvik, L. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., Downing, J. A., Middelburg, J. J., & Melack, J. (2007). Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems, 10(1), 171–184. https://doi.org/10.1007/S10021-006-9013-8

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

Fichot, C. G., Benner, R., Kaiser, K., Shen, Y., Amon, R. M. W., Ogawa, H., & Lu, C. J. (2016). Predicting dissolved lignin phenol concentrations in the coastal ocean from chromophoric dissolved organic matter (CDOM) absorption coefficients. Frontiers in Marine Science, 3(FEB). https://doi.org/10.3389/fmars.2016.00007

Fischer, K., & Höffler, S. (2021). RP-HPLC Analysis of Phenolic Lignin Monomers with DAD and Coulometric Array Detection. Journal of Chromatographic Science, 59(4), 312–323. https://doi.org/10.1093/CHROMSCI/BMAA105

Freeman, C., Evans, C. D., Monteith, D. T., Reynolds, B., & Fenner, N. (2001). Export of organic carbon from peat soils. Nature, 412(6849), 785. https://doi.org/10.1038/35090628

Gardiner, D. T., & Miller, R. W. (2004). Soils In Our Environment (10th ed.). Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA.

Harvey, H. R., & Mannino, A. (2001). The chemical composition and cycling of particulate and macromolecular dissolved organic matter in temperate estuaries as revealed by molecular organic tracers. Organic Geochemistry, 32(4), 527–542. https://doi.org/10.1016/S0146-6380(00)00193-5

Hedges, J. I., & Ertel, J. R. (1982). Characterization of Lignin by Gas Capillary Chromatography of Cupric Oxide Oxidation Products. Analytical Chemistry, 54(2), 174–178. https://doi.org/10.1021/AC00239A007

Hedges, J. I., & Mann, D. C. (1979). The characterization of plant tissues by their lignin oxidation products. Geochimica et Cosmochimica Acta, 43(11), 1803–1807. https://doi.org/10.1016/0016-7037(79)90028-0

Hernes, P. J., & Benner, R. (2002). Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean. In Deep-Sea Research I (Vol. 49).

Hernes, P. J., & Benner, R. (2003). Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. Journal of Geophysical Research: Oceans, 108(9). https://doi.org/10.1029/2002jc001421

Ingalls, A. E., Ellis, E. E., Santos, G. M., McDuffee, K. E., Truxal, L., Keil, R. G., & Druffel, E. R. M. (2010). HPLC purification of higher plant-dervied lignin phenols for compound specific radiocarbon analysis. Analytical Chemistry, 82(21), 8931–8938. https://doi.org/10.1021/ac1016584

Johnsen, L. G., Amigo, J. M., Skov, T., & Bro, R. (2014). Automated resolution of overlapping peaks in chromatographic data. Journal of Chemometrics, 28(2), 71–82. https://doi.org/10.1002/CEM.2575

Kaiser, K., & Benner, R. (2012). Characterization of lignin by gas chromatography and mass spectrometry using a simplified CuO oxidation method. Analytical Chemistry, 84(1), 459–464. https://doi.org/10.1021/ac202004r

Kaiser, K., Benner, R., & Amon, R. M. W. (2017). The fate of terrigenous dissolved organic carbon on the Eurasian shelves and export to the North Atlantic. Journal of Geophysical Research: Oceans, 122(1), 4–22. https://doi.org/10.1002/2016JC012380

Lewis, N. G., & Yamamoto, E. (1990). LIGNIN: OCCURRENCE, BIOGENESIS AND BIODEGRADATION. Annu. Rev. Plant Physiol. Plant Mol. BioI, 41, 455–496.

Lobbes, J. M., Fitznar, H. P., & Kattner, G. (1999). High-performance liquid chromatography of lignin-derived phenols in environmental samples with diode array detection. Analytical Chemistry, 71(15), 3008–3012. https://doi.org/10.1021/ac981175i

Lobbes, J. M., Fitznar, H. P., & Kattner, G. (2000). Biogeochemical characteristics of dissolved and particulate organic matter in Russian rivers entering the Arctic Ocean. Geochimica et Cosmochimica Acta, 64(17), 2973–2983. https://doi.org/10.1016/S0016-7037(00)00409-9

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology, 2(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Louchouam, P., Opsahl, S., & Benner, R. (2000). Isolation and Quantification of Dissolved Lignin from Natural Waters Using Solid-Phase Extraction and GC/MS. Analytical Chemistry, 72(13), 2780–2787. https://doi.org/10.1021/AC9912552

Louchouarn, P., Amon, R. M. W., Duan, S., Pondell, C., Seward, S. M., & White, N. (2010). Analysis of lignin-derived phenols in standard reference materials and ocean dissolved organic matter by gas chromatography/tandem mass spectrometry. Marine Chemistry, 118(1–2), 85–97. https://doi.org/10.1016/j.marchem.2009.11.003

Mann, P. J., Spencer, R. G. M., Hernes, P. J., Six, J., Aiken, G. R., Tank, S. E., McClelland, J. W., Butler, K. D., Dyda, R. Y., & Holmes, R. M. (2016). Pan-arctic trends in terrestrial dissolved organic matter from optical measurements. Frontiers in Earth Science, 4. https://doi.org/10.3389/feart.2016.00025

Monties, B., & Fukushima, K. (2001). Occurrence, Function and Biosynthesis of Lignins. Biopolymers Online. https://doi.org/10.1002/3527600035.BPOL1001

Murphy, K. R., Stedmon, C. A., Graeber, D., & Bro, R. (2013). Fluorescence spectroscopy and multi-way techniques. PARAFAC. Analytical Methods, 5(23), 6557–6566. https://doi.org/10.1039/C3AY41160E

Opsahl, S., & Benner, R. (1997). Distribution and cycling of terrigenous dissolved organic matter in the ocean. Nature 1997 386:6624, 386(6624), 480–482. https://doi.org/10.1038/386480a0

Opsahl, S., & Benner, R. (1998). Photochemical reactivity of dissolved lignin in river and ocean waters. Limnology and Oceanography, 43(6), 1297–1304. https://doi.org/10.4319/LO.1998.43.6.1297

Osburn, C. L., Boyd, T. J., Montgomery, M. T., Bianchi, T. S., Coffin, R. B., & Paerl, H. W. (2016). Optical proxies for terrestrial dissolved organic matter in estuaries and coastal waters. Frontiers in Marine Science, 2(JAN), 127. https://doi.org/10.3389/FMARS.2015.00127/BIBTEX

Osburn, C. L., & Stedmon, C. A. (2011). Linking the chemical and optical properties of dissolved organic matter in the Baltic-North Sea transition zone to differentiate three allochthonous inputs. Marine Chemistry, 126(1–4), 281–294. https://doi.org/10.1016/j.marchem.2011.06.007

Petersen, M. A., & Bro, R. (2018). PARADISe - a ground-breaking tool to treat complex GC-MS datasets. In B. Siegmund & E. Leitner (Eds.), Flavour Science: Proceedings of the XV Weurman Flavour Research Symposium (Vol. 350, pp. 421–426). Technischen Universität Graz.

Quinton, J. N., Govers, G., van Oost, K., & Bardgett, R. D. (2010). The impact of agricultural soil erosion on biogeochemical cycling. Nature Geoscience, 3(5), 311–314. https://doi.org/10.1038/NGEO838

Rayens, W. S., & Mitchell, B. C. (1997). Two-factor degeneracies and a stabilization of PARAFAC. Chemometrics and Intelligent Laboratory Systems, 38(2), 173–181. https://doi.org/10.1016/S0169-7439(97)00033-6

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., Laruelle, G. G., Lauerwald, R., Luyssaert, S., Andersson, A. J., Arndt, S., Arnosti, C., Borges, A. v., Dale, A. W., Gallego-Sala, A., Goddéris, Y., Goossens, N., Hartmann, J., Heinze, C., …

Thullner, M. (2013). Anthropogenic perturbation of the carbon fluxes from land to ocean. Nature Geoscience, 6(8), 597–607. https://doi.org/10.1038/NGEO1830

Reuter, H., Gensel, J., Elvert, M., & Zak, D. (2017). Direct Analysis of Lignin Phenols in Freshwater Dissolved Organic Matter. Analytical Chemistry, 89(24), 13449–13457. https://doi.org/10.1021/acs.analchem.7b03729

Shrivastava, A., & Gupta, V. (2011). Methods for the determination of limit of detection and limit of quantitation of the analytical methods. Chronicles of Young Scientists, 2(1), 21. https://doi.org/10.4103/2229-5186.79345

Skov, T., & Bro, R. (2008). Solving fundamental problems in chromatographic analysis. Analytical and Bioanalytical Chemistry, 390(1), 281–285. https://doi.org/10.1007/S00216-007-1618-Z

Spencer, R. G. M., Mann, P. J., Dittmar, T., Eglinton, T. I., McIntyre, C., Holmes, R. M., Zimov, N., & Stubbins, A. (2015). Detecting the signature of permafrost thaw in Arctic rivers. Geophysical Research Letters, 42(8), 2830–2835. https://doi.org/10.1002/2015GL063498

Stedmon, C. A., Markager, S., & Bro, R. (2003). Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Marine Chemistry, 82(3–4), 239–254. https://doi.org/10.1016/S0304-4203(03)00072-0

Steinberg, S., Venkatesan, M. I., & Kaplan, I. R. (1984). Analysis of the products of the oxidation of lignin by CuO in biological and geological samples by reversed-phase high-performance liquid chromatography. Journal of Chromatography A, 298(C), 427–434. https://doi.org/10.1016/S0021-9673(01)92740-2

Tucker, L. R. (1951). A method for synthesis of factor analysis studie. Psychometrika, 16(3), 121–136. https://doi.org/10.21236/AD0047524

Vonk, J. E., Mann, P. J., Davydov, S., Davydova, A., Spencer, R. G. M., Schade, J., Sobczak, W. v., Zimov, N., Zimov, S., Bulygina, E., Eglinton, T. I., & Holmes, R. M. (2013). High biolability of ancient permafrost carbon upon thaw. Geophysical Research Letters, 40(11), 2689–2693. https://doi.org/10.1002/grl.50348

Vonk, J. E., Sanchez-Garca, L., van Dongen, B. E., Alling, V., Kosmach, D., Charkin, A., Semiletov, I. P., Dudarev, O. v., Shakhova, N., Roos, P., Eglinton, T. I., Andersson, A., &

Gustafsson, A. (2012). Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia. Nature, 489(7414), 137–140. https://doi.org/10.1038/NATURE11392

Walker, S. A., Amon, R. M. W., Stedmon, C., Duan, S., & Louchouarn, P. (2009). The use of PARAFAC modeling to trace terrestrial dissolved organic matter and fingerprint water masses in coastal Canadian Arctic surface waters. Journal of Geophysical Research: Biogeosciences, 114(4). https://doi.org/10.1029/2009JG000990

Wild, B., Andersson, A., Bröder, L., Vonk, J., Hugelius, G., McClelland, J. W., Song, W., Raymond, P. A., & Gustafsson, Ö. (2019). Rivers across the Siberian Arctic unearth the patterns of carbon release from thawing permafrost. Proceedings of the National Academy of Sciences of the United States of America, 116(21), 10280–10285. https://doi.org/https://doi.org/10.1073/pnas.1811797116

Yan, G., & Kaiser, K. (2018a). Ultralow Sample Volume Cupric Sulfate Oxidation Method for the Analysis of Dissolved Lignin. Analytical Chemistry, 90(15), 9289–9295. https://doi.org/10.1021/acs.analchem.8b01867

Yan, G., & Kaiser, K. (2018b). A rapid and sensitive method for the analysis of lignin phenols in environmental samples using ultra-high performance liquid chromatography-electrospray ionization-tandem mass spectrometry with multiple reaction monitoring. Analytica Chimica Acta, 1023, 74–80. https://doi.org/10.1016/j.aca.2018.03.054

Yu, H., Bro, R., & Gallagher, N. B. (2023). PARASIAS: A new method for analyzing higher-order tensors with shifting profiles. Analytica Chimica Acta, 1238, 339848. https://doi.org/10.1016/J.ACA.2022.339848

Acknowledgements

This study was financed from the Independent Research Fund Denmark Grant No. 9040-00266B awarded to C.A.S.. We acknowledge Kaj & Hermille Fund, who supported the visit to NCSU for A.D.B., which helped to expand on lignin analysis using gas chromatography and mass spectrometry. We want to thank Professor Torkel Gissel Nielsen and former Ph.D. student Christian Mathias Rohde Kiær for providing the Kattegat samples. We also want to thank Associate Professor John Kominoski for supplying the FCE samples. This material is based upon work supported by the National Science Foundation through the Florida Coastal

Everglades Long-Term Ecological Research program under Grant No. DEB-1832229. Finally, we would like to thank the entire research team and the crew aboard R/V Dana during the two cruises for helping out with the sampling.

Supplymentary Information: Lignin phenol quantification from machine learning-assisted decomposition of liquid chromatography-absorbance spectroscopy data.

Anders Dalhoff Bruhn[1], Urban Wünsch[1], Christopher L. Osburn[2], Jacob C. Rudolph[2,3] and Colin A. Stedmon[1].

[1]National Institute of Aquatic Resources, Technical University of Denmark, Denmark.

[2]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, USA.

[3]Department of Environmental Biology, Smithsonian Environmental Research Center, Edgewater, MD

Table S1: The concentration (nM) of lignin phenols in each of the calibration standards.

	Lignin phenol											
Standard no.	**PAD**	**VAD**	**SAD**	**PAL**	**CAD**	**PON**	**VAL**	**SAL**	**FAD**	**VON**	**SON**	**CIN**
1	25000	10000	5000	2500	1000	500	200	50	25	10	5	1
2	10000	5000	2500	1000	500	200	50	25	10	5	2	25000
3	5000	2500	1000	500	200	50	25	10	5	2	1	10000
4	2500	1000	500	200	50	25	10	5	2	1	25000	5000
5	1000	500	200	50	25	10	5	2	1	25000	10000	2500
6	500	200	50	25	10	5	2	1	25000	10000	5000	1000
7	200	50	25	10	5	2	1	25000	10000	5000	2500	500
8	50	25	10	5	2	1	25000	10000	5000	2500	1000	200
9	25	10	5	2	1	25000	10000	5000	2500	1000	500	50
10	10	5	2	1	25000	10000	5000	2500	1000	500	200	25
11	5	2	1	25000	10000	5000	2500	1000	500	200	50	10
12	2	1	25000	10000	5000	2500	1000	500	200	50	25	5
13	1	25000	10000	5000	2500	1000	500	200	50	25	10	2
14	25	2	1	10	25	10	1	2	25	2	1	25
15	1	5	10	25	5	25	2	1	5	25	10	1
16	10	25	5	5	1	2	10	25	10	10	5	10
17	2	1	2	1	2	1	5	10	2	1	2	2
18	5	10	25	2	10	5	25	5	1	5	25	5

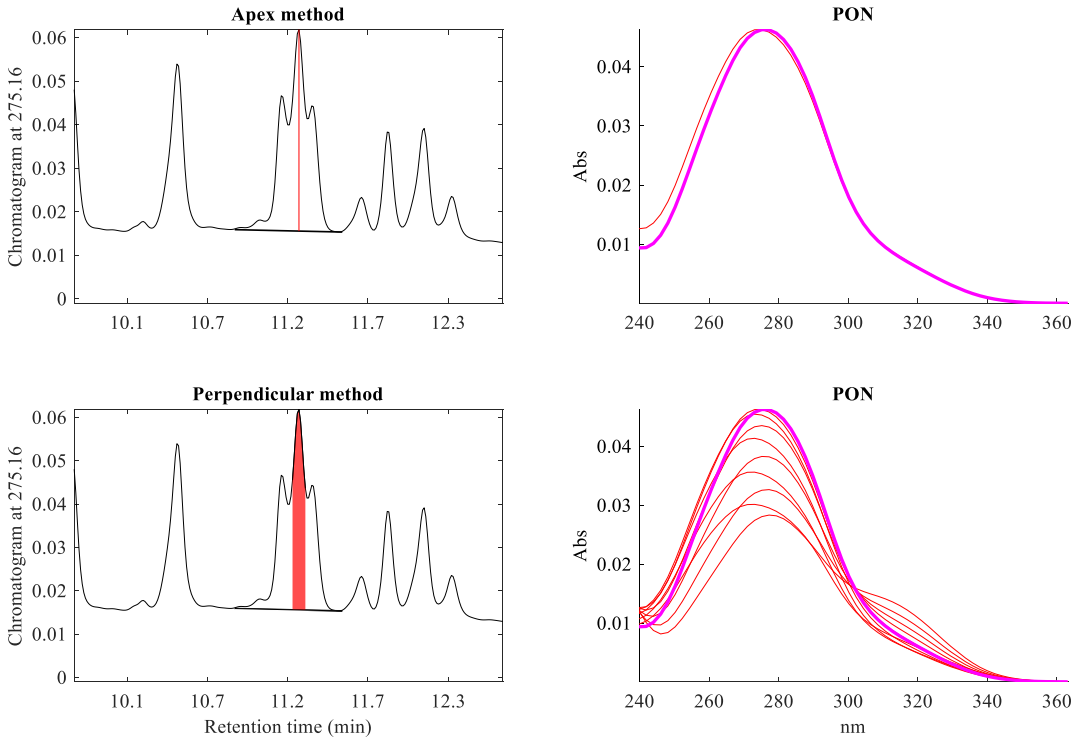


Figure S1: Example of the two manual integration methods on a wavelength-specific chromatogram extracted from the measurement of a spiked sample for the quantification of PON. The spectra across the integrated peaks (red line(s)) are compared to the pure spectrum (magenta) of the lignin phenol.

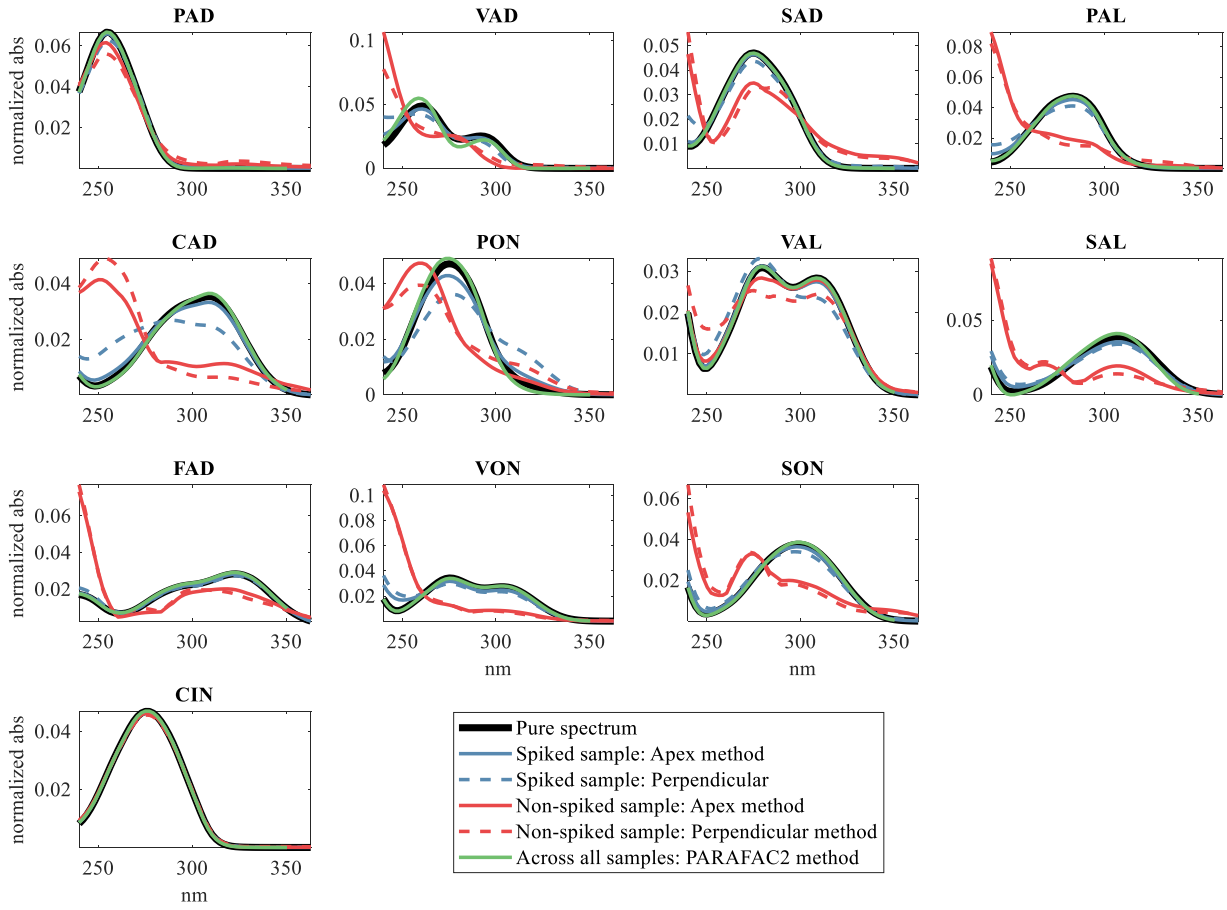


Figure S2: Lignin phenol spectra extracted by the three different quantification methods (PARAFAC2, apex method, perpendicular drop method) and the pure spectra extracted from their most concentrated calibration standards. The spectra are normalized for comparison among the methods and for use in the TCC test. For the perpendicular drop method, the extracted spectra shown above, are the mean of all spectra at each retention point across the integrated peaks.

Assessment of effect of DOC concentration on lignin phenols and ratios (larger dataset)

Lignin phenol data from 57 seawater samples taken across the Fram Strait (unpublished data) were examined in the same manner as the Kattegat samples. The concentration of DOC in the reaction vessel in the 57 samples ranged from 7-165 mM. Similar to the Kattegat samples, no visual correlation was seen between the parameters and the DOC concentration (see Figure S3 & S4). Dividing the samples into two groups, using the mean DOC concentration (60 mM) as the borderline to divide them, did not show any significant (p-values=0.30-0.92, see Figure S3) difference in the final environmental concentration of any lignin phenols, indicating no trend with increased DOC concentration in the reaction vessel. Also for the diagenetic ratios, no significant difference was demonstrated between the two groups (p-values between 0.08-0.76, see Figure S4) which as well indicate no trend in ratios with increasing DOC concentration in the reaction vessel.

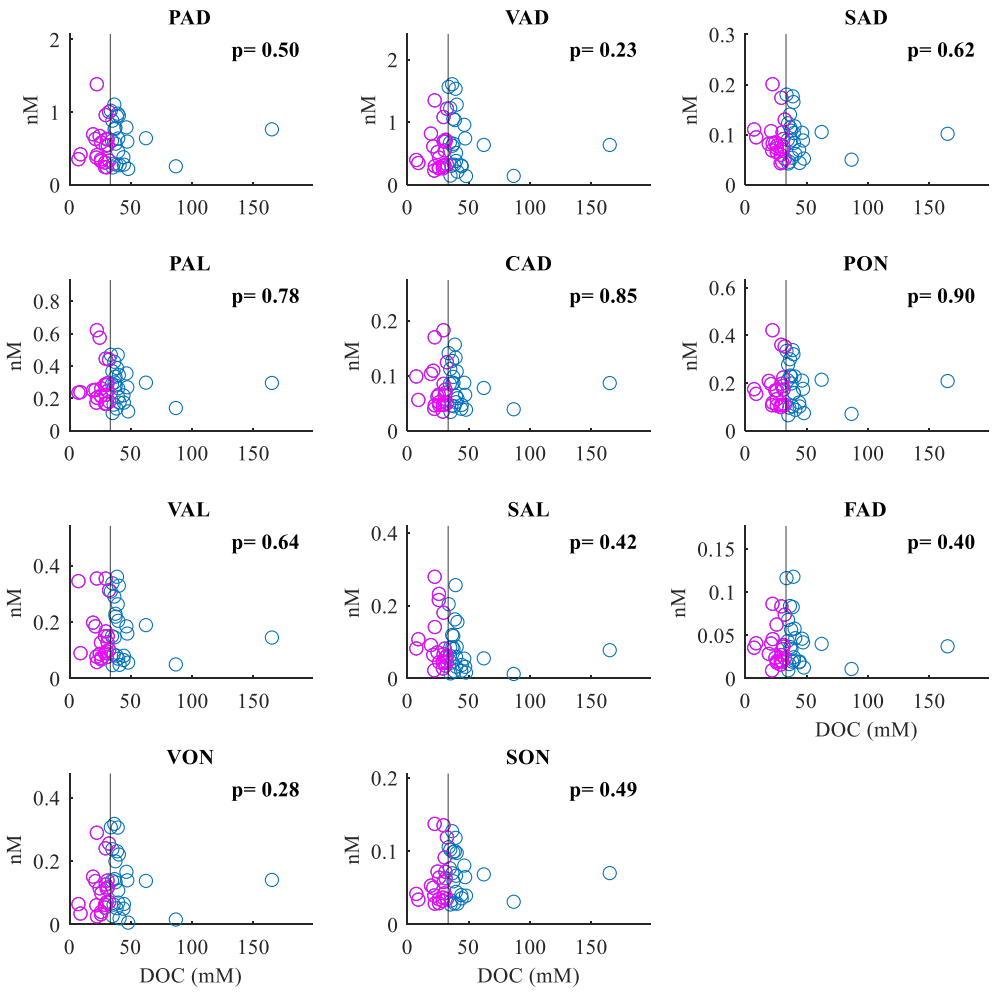


Figure S3:Lignin phenol in their environmental concentrations (nM) plotted against the DOC concentration (mM) in the reaction vessel for 57 samples collected in Fram Strait (unpublished data). The black line indicates the mean DOC concentration used to divide the samples into two groups (purple and blue) for statistical analysis.

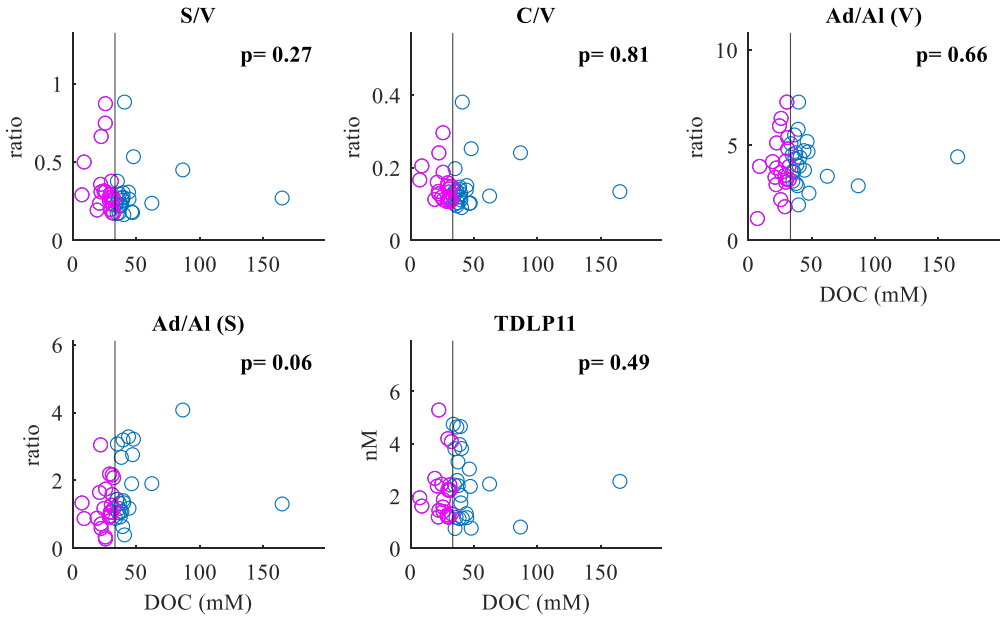


Figure S4:Diagenetic ratios (unitless) and environmental TDLP11 (nM) plotted against the DOC concentration (mM) in the reaction vessel for 57 samples collected in Fram Strait (unpublished data). The black line indicate the mean DOC concentration used to divide the samples into two groups (purple and blue) used for statistical analysis.

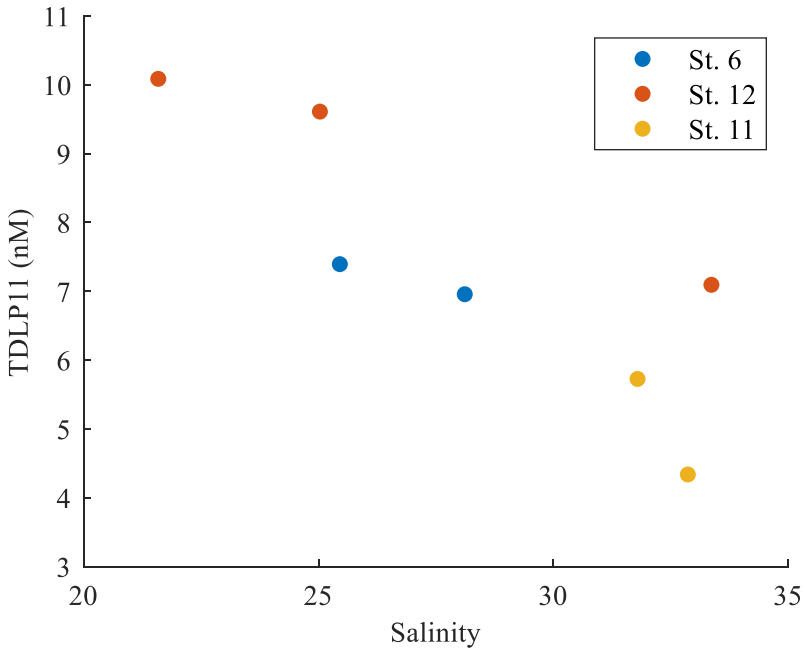


Figure S5: Relationship between TDLP11 (nM) and salinity for the samples taken at the three stations in the Kattegat.

Paper C

Dissolved lignin phenols across the Fram Strait: Towards in-situ measurements from fluorescence

Anders Dalhoff Bruhn, Urban Wünsch, Christopher L. Osburn, Rafael Gonçalves-Araujo, Mats A. Granskog**,** and Colin A. Stedmon.

This paper is in the last stage of preparation and will be submitted to Limonology & Oceanography.

Dissolved lignin phenols across the Fram Strait: Towards in-situ measurements from fluorescence.

Anders Dalhoff Bruhn[1], Urban Wünsch[1], Christopher L. Osburn[2], Rafael Gonçalves-Araujo[1], Mats A. Granskog[3], Colin A. Stedmon[1].

[1]National Institute of Aquatic Resources, Technical University of Denmark, Denmark.

[2] Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, USA.

[3] Norwegian Polar Institute, Fram Centre, Tromsø, Norway.

Correspondence to: adbj@aqua.dtu.dk

Keywords: Lignin phenols, fluorescence, N-PLS, Fram Strait, terrestrial DOM, PLS

ABSTRACT

The Fram Strait is the major gateway between the Arctic and the Atlantic Ocean. As Atlantic water (AW) enters the Arctic Ocean, it accumulates freshwater and terrestrial dissolved organic matter (tDOM) from major Arctic rivers. Enriched in tDOM, the Polar surface water (PSW) returns to Fram Strait. Lignin is a tDOM biomarker which can be used to characterize its origin and passage through the Arctic Ocean. The goal was therefore to characterise lignin in the Fram Strait and link it to the fluorescence properties of DOM. Lignin phenol concentrations were quantified using High-Pressure Liquid Chromatography coupled with absorbance detection. The measurements showed total dissolved lignin phenol (TDLP11) concentrations across the Fram Strait between 0.8 and 4.8 nM, where highest concentrations were found in PSW. Lignin phenol composition suggested that AW predominantly carries angiosperm derived material in the surface waters, whereas PSW carries Arctic river-derived material. Ratios of lignin phenols suggested that DOM in PSW possibly consists of two different terrestrial sources. A N-way Partial Least Square (N-PLS) model was trained to predict lignin parameters (r^2=0.24-0.76) from excitation-emission-matrices (EEMs) of seawater from the Fram Strait. The N-PLS model was applied to 327 EEMs from samples taken across the Fram Strait and reflected expected TDLP11 concentrations. A second N-PLS model was trained using emission spectra from only four excitation wavelengths and showed equal prediction power for the lignin parameters (r^2 +/- 0.05). This implies that lignin phenols can be estimated using measurements from in-situ sensors already commercially available. Such capabilities pave the way to trace distributions of terrestrial DOM with higher spatial resolution and on longer time scales in the near future.

INTRODUCTION

The Fram Strait is one of the major gateways for water exchange between the Arctic and the Atlantic Ocean. Associated with the exchange in water masses there is an exchange in dissolved organic matter (DOM) with contrasting origin and properties. The two major currents in the Fram Strait are the northbound West Spitsbergen Current (WSC) and the southbound East Greenland Current (EGC). The WSC transports warm and saline Atlantic water (AW) northwards in the eastern side of the Fram Strait, whereas the EGC transports stratified Arctic water southwards in the western side of the Fram Strait. The EGC flows over the East Greenland shelf (10-17 °W at 78°50' N) and along East Greenland slope (6-9 °W at 78°50' N) and consists of three larger water masses, based on salinity and temperature, herein Polar surface waters (PSW), Recirculated AW (RAW) and Modified AW (MAW) (Amon et al., 2003).

DOM supplied from the Atlantic consists of an aged oceanic signal of processed marine DOM from surface plankton production and refractory terrestrial DOM. In contrast, surface waters in the Arctic outflow carry a great amount of dissolved organic carbon (DOC; 46.8 (±6.2) Tg C yr^{-1} ; Gonçalves-Araujo et al., 2020) with a dominant signature of terrestrial DOM (tDOM) (Amon et al., 2003; Gonçalves-Araujo et al., 2016, 2020b; Granskog et al., 2012). This signature has its origins in the large catchments of Siberian rivers and to a lesser extent North American rivers (Amon et al., 2003; Stedmon et al., 2011, 2021). Since the Arctic Ocean represents only 1% of the global ocean volume but receives 10% of the global river runoff (McClelland et al., 2012), it has a greater average concentration of DOC relative to other ocean basins and 14-24% of the DOC in Arctic Ocean is found to be terrestrial (Benner et al., 2005; Opsahl et al., 1999). Due to accelerating global warming in the Arctic, a substantial increase in shelf-derived material in the Arctic Ocean over the past decade has been observed (Kipp et al., 2018). Additionally, discharge from Arctic rivers is estimated to increase with 14% by the end of 21st century (Haine et al., 2015). The amount of tDOM exported to the Arctic Ocean is therefore expected increase and the fate and impact of this increased supply on marine ecosystems is unclear. Due to its light absorption properties, it can reduce the photic depth (Pavlov et al., 2015) and contribute to increased solar heating of surface waters (Granskog et al., 2015; Hill, 2008; Pegau, 2002). If a fraction of it is available to microbes it can result in a change in the competition for nutrients between heterotrophs and autotrophs (Thingstad et al., 2008). The ultimate fate for tDOM is either microbial or photochemical mineralisation (Bruhn et al., 2021; Cory et al., 2013, 2014; Sipler et al., 2017) or export to the deep sea during bottom

water formation Mineralisation in surface waters can potentially contribute to increasing atmospheric CO_2 concentrations, whereas sequestration to the deep sea does not.

Lignin is a phenolic biopolymer which is only found in vascular plants and is very resistant towards microbial degradation. Therefore, it is an excellent tracer of tDOM in the oceans. Lignin is however prone to photodegradation (Benner & Kaiser, 2011; Spencer et al., 2009). Most of the degradation takes place in the coastal zone and the lesser susceptible portion of lignin that reaches the open ocean can persist for 100 years (Hernes & Benner, 2002, 2003; Opsahl & Benner, 1998). Lignin quantification is traditionally accomplished by oxidation of the biopolymer into its constituent phenolic monomers, also known as lignin phenols (Hedges & Ertel, 1982; Lobbes et al., 1999; Yan & Kaiser, 2018). Their combined concentration, total dissolved lignin phenols (TDLP), are in the nanomolar or picomolar range in seawater. The high tDOM content of the Arctic Ocean is reflected in TDLP concentrations, which have been measured to vary between 0.3-17 nM, with highest concentrations in the East Siberian Sea (TDLP9=8-17 nM), intermediate concentrations in PSW (TDLP6=1.6 nM), and lowest concentrations in deep waters (TDLP6=0.25 nM,) (Benner et al., 2005; Kaiser et al., 2017). In contrast, the TDLP concentration in the Pacific and the Atlantic Ocean is much lower, reaching up to 60 and 154 pM respectively (Opsahl & Benner, 1997). Siberian rivers, entering the Arctic Ocean, have higher TDLP concentrations, varying between 34-238 nM depending on the river (Amon et al., 2012; Lobbes et al., 2000).

The composition of lignin phenols can be used as an indicator of source and exposure to degradation processes (Hedges et al., 1988; Hedges & Mann, 1979; Opsahl & Benner, 1995, 1998). The ratio of syringyl to vanillyl (S/V) and cinnamyl to vanillyl (C/V) phenols can be used to distinguish between different sources (Amon et al., 2003; Lobbes et al., 2000). In contrast to S and C phenols, p-hydroxyl phenols (P) are widely distributed among all vascular plants and a change in P/V is not as specific to source material as the S/V and C/V ratios (Fichot et al., 2016). The P/V can however be used to indicate photodegradation (Benner & Kaiser, 2011; Kaiser et al., 2017) as the ratio increase with longer exposure to sunlight.

Although lignin measurements offer an indisputable tDOM tracer, quantification involves laborious and time-consuming procedures, such as solid phase extraction (SPE), chemical oxidation, chromatography, and identification of a suite of specific lignin phenols (Dittmar et al., 2008; Hedges & Ertel, 1982; Kaiser & Benner, 2012; Lobbes et al., 1999; Yan & Kaiser, 2018). Developing proxies to other DOM measurements could therefore beneficial increase the

spatio- temporal resolution of surveys (Spencer et al 2008; Hernes et al 2009). Fichot et al. (2012, 2016) has shown a strong correlation between TDLP concentration and CDOM absorption in coastal waters of the Gulf of Mexico. In the same region a later study by Yamashita et al. (2015) found strong correlations between DOM fluorescence and TDLP. Hernes et al. (2009) applied bilinear Partial Least Square (PLS) of spectral fluorescence (excitation-emission-matrices, EEMs) to predict TDLP and diagenetic ratios in river waters. Although there is a contribution of the absorption and fluorescence of individual lignin phenols to the combined DOM absorption and fluorescence (non lignin and lignin derived material) the contribution from other substances is very high. Therefore these relationships represent local empirical relationships which if robust for a particular region offer valuable approach for scaling up the coverage of lignin estimates.

Here we expand on the progress so far to study the relationship between lignin content and characteristics and DOM fluorescence in Arctic Ocean waters, where fluorescence intensities and lignin phenol concentrations are lower than in the earlier studies. For this we apply an approach named N-way Partial Least Squares (N-PLS) which retains the trilinear nature of the EEM data, leading to increased interpretability and better prediction compared to bilinear PLS (Bro, 1996, 1998). Quantifying specific compounds by applying N-PLS to EEMs has already been successfully performed by multiple studies in other environments than the ocean (Bai et al., 2018; Kumar & Mishra, 2012; Lin et al., 2022; Matero et al., 2010).

If successful this approach provides an affordable and rapid analytical technique to estimate and monitor the distribution and origin of tDOM in water masses flowing in and out from the Arctic Ocean and better follow the fate of terrestrial carbon. Ultimately, such an approach can pave the way for development of in-situ fluorescence sensors for tracing terrestrial DOM in the ocean.

MATERIALS & PROCEDURES

Sample collection

Seawater samples were taken during the Norwegian Polar Institute's (NPI) Fram Strait cruise in 2021 onboard R/V Kronprins Haakon as part of the NoTAC project ("Novel Tracers of Arctic Carbon and water exchange in the Fram Strait"). The cruise was part of the Fram Strait Arctic Outflow Observatory, whose primary purpose is to service the moorings in the in the

western Fram Strait and make a hydrographic survey across the strait.The sampling stations span across the Fram Strait at 78°50' N. Seawater samples for DOM fluorescence were taken at 21 stations (black dots in Figure 1A), while samples for lignin were collected at seven stations (6°, 8°, 10°, 12°, 15°, 17° W and 7.5° E). Samples for fluorescent DOM (FDOM) were collected at fixed depths (1, 5, 15, 25, 50, 100, 150, 200, 250, 400, 500 m depth; see black dots in Figure 1D), while samples for lignin were collected at varying depths, down to 300 m (see Figure 2). Seawater was collected using Niskin bottles mounted on a water sampling carousel equipped with a SBE911plus conductivity-temperature-depth (CTD) system (Seabird Scientific, USA). A total of 53 seawater samples were taken for lignin analysis, while 327 samples were collected for FDOM measurement.

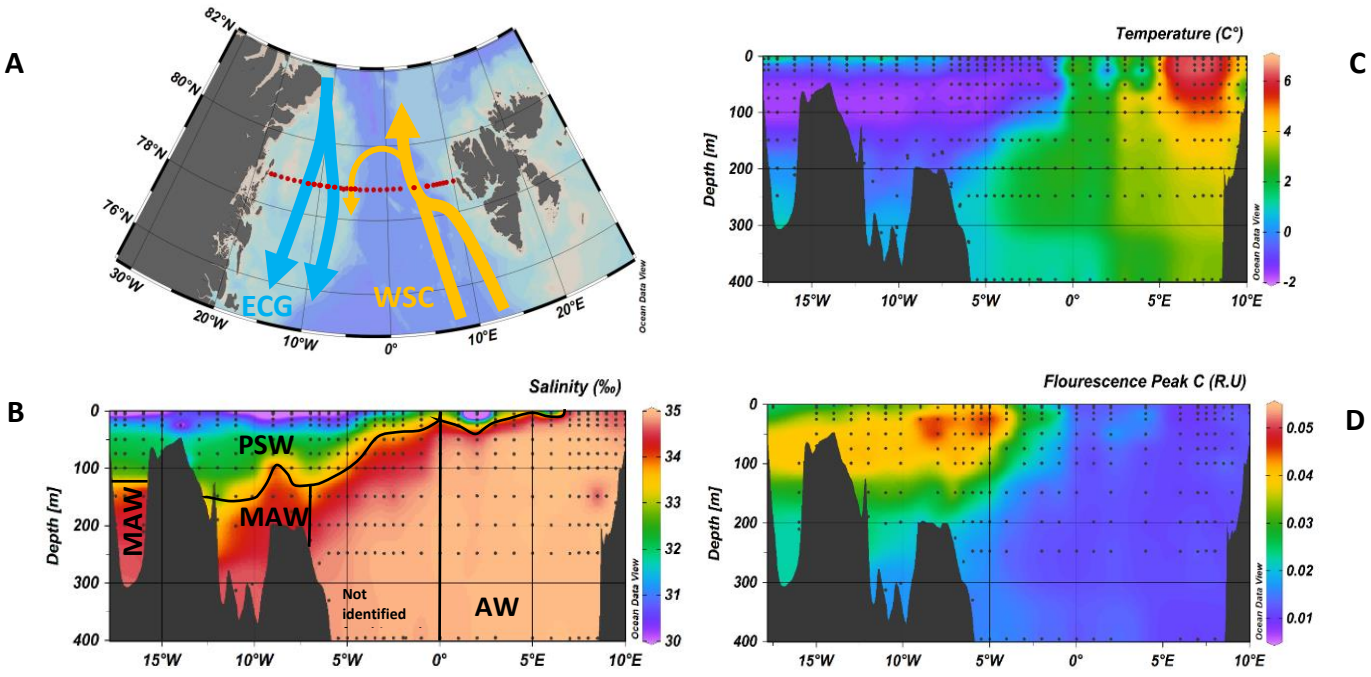


Figure 1: A) Location of the Fram Strait transect stations in 2021 and the ocean currents travelling through; B-D) section plots of B) practical salinity, including defined water masses (see Table 1 for oceanographic definitons), C) temperature and D) DOM fluorescence (peak C: excitation 350, emission 450 (Coble, 2007)), respectively. The black dots in B-D indicate the water sampling depths for FDOM samples. Samples from greater depth (n=26) are not shown and are from 10 m above the seafloor.

Samples for FDOM were filtered directly from the Niskin bottles through a Millipore Opticap 0.2 µm filter cartridge into 40 mL combusted brown acid-washed, pre-combusted glass vials. For lignin phenols, the sampling volume varied depending on water masses and the expected

DOM content. For samples located within the Arctic outflow (depths above 150 m for the western stations) 5 L was collected, while 10 L was taken for the remainder depths and stations. Lignin samples were filtered through a 0.2 µm cartridge filter (Polyethersulfone Membrane Capsule Filter, Sterlitech Inc.) immediately after collection, then acidified to approximately pH 2 using hydrochloric acid (HCl) and stored in a dark and cold room (4 °C) until extraction of DOM was performed (within 1-2 days onboard the ship for the first 38 samples, one month later for the remaining 20 samples).

Measurement of DOM fluorescence

DOM fluorescence of the filtered samples was characterized in a 1 cm quartz cuvette using a Horiba AquaLog fluorometer onboard the research vessel. Emission was measured at wavelengths between 280–620 nm and excitation was measured between 240–450 nm with increment of five nm (8 seconds integration time). A water reference (Starna Scientific) was measured each day, prior to seawater samples, for Raman calibration and was afterwards used as a blank for the rest of the measurements on the same day. Fluorescence counts were converted into Raman units (R.U) applying Raman calibration (Lawaetz & Stedmon, 2009) and first and second order Raman and Rayleigh scatter were removed.

Lignin extraction and oxidation

Solid phase extraction was performed on the filtered and acidified seawater using Priority PolLutant (PPL) cartridges (200 mg sorbent, 3 mL cartridge capacity, Bond Elut, Agilent). The PPL cartridges were cleaned with 3 mL methanol (High-Pressure Liquid Chromatography (HPLC) grade) and conditioned with 6 mL of acidified ultrapure water (pH 2, HCl) prior to extraction. After seawater extraction, the PPL cartridges were dried using a vacuum hand pump and stored at 4 °C (in pre-combusted and sealed brown vials) until further processing after the cruise (approximately four months later).

The DOM absorbed onto the PPL cartridge was eluted with 3 mL of methanol. A 0.6 mL subsample was taken from the methanol and transferred to a custom-designed steel cylinder reaction vessel (850 uL) where the DOM was dried down under nitrogen gas and 45 µL of 10 mM $CuSO_4$, 40 µL of 0.2 M ascorbic acid (antioxidant to avoid over-oxidation) and 748 µL of 1.1 M NaOH were added to the reaction vessel. The reaction vessel was then sealed and placed in an oven for two hours at 150 °C. After cooling the vessel down under running water, the

reaction solution was transferred to a glass vial where an internal standard (17 µL of 0.25 mM cinnamic acid to give a final concentration of 21.25 µM) was added and the whole solution was acidified to pH 2 with 85 µL of H_2SO_4. The reaction solution was then pipetted onto a preconditioned Hydrophilic Lipophilic Balanced (HLB) cartridge (200 mg sorbent) where 3 x 300 µL of a 20% (v/v) methanol / water mixture was added to remove inorganic and polar compounds in the solution. The HLB cartridge was then dried using nitrogen gas. The remaining DOM was then eluted using a 30% (v/v) methanol / methyl acetate. The methanol / methyl acetate solution was collected in a small glass vial and then dried using nitrogen gas. The dried sample was reconstituted in 190 µL ultrapure water with pH adjusted to 2.5 with 10 µL of 0.14 M H_3PO_4. The 200 µL aqueous sample containing the oxidized DOM was then transferred to a HPLC vial with a 250 µL insert vial and stored at -18 °C until analysis (approximately one month later).

Lignin analysis

Phenolic monomers were used as analytical calibration standards. 4-hydroxybenzoic acid (PAD), 4-hydroxy-cinnamic acid (CAD) and cinnamic acid (CIN) were obtained from Merck. 4-hydroxybenzaldehyd (PAL), syringic acid (SAD), vanillin (VAN), syringaldehyde (SAL) were obtained from Alfa Aeasar. Acetovanillone (VON), ferulic acid (FAD) and acetosyringone (SON) were obtained from Acros Organics. 4-hydroxyacetophenone (PON) and vanillic acid (VAD) were obtained from Sigma Aldrich and TCI Europe, respectively. Stock solutions with a concentration of 1.25 mM in ultrapure water were made for each of the 12 phenolic monomers (11 lignin phenols and 1 internal standard). The stock solutions were stored at -18 °C until further use.

Calibration standards were created by diluting the stock solutions in ultrapure water. The calibration curves consisted of 13 concentrations ranging between 1 nM and 25 µM. A reference standard mixture (20 µM of each lignin phenol) was made by diluting the stock solutions in ultrapure water and was later used to track large retention time shifts during analysis. An additional reference was made using Suwannee River natural organic matter reference material (SRNOM, Cat. Num. 2R101N) purchased from the International Humic Substances Society, isolated in 2012 using reverse osmosis (Green et al., 2015) and was later used to track any change caused by the matrix effect.

Separation of the oxidized DOM was performed by HPLC on a Shidmadzu Nexera X2 HPLC system equipped with LC-20AB pumps using a C18 column (Poroshell 120 EC-C18 4.6 x 100 mm, 2.7 µm pore diameter, Agilent) with a guard column attached. The two mobile phases used for the liquid chromatography consisted of 7.4 mM phosphoric acid in ultrapure water (mobile phase A; pH 2.4) and pure acetonitrile (mobile phase B). The total runtime for each sample was 40 minutes. Diode-Array Detection (DAD; SPD-M30A, Shimadzu) was used for quantification and identification of targeted lignin phenols was therefore based of their absorbance spectrum.

The chromatograms were analysed using Parallel Factor Analysis 2 (PARAFAC2) which isolates individual elution profiles and spectra of the phenolic monomers and is capable of separating highly co-eluting peaks (Bruhn et al. (in review)). The concentration of each phenolic monomer was calculated using the relationship between the PARAFAC2 score and concentration determined from the calibration curves. The environmental concentration (nM) of each lignin phenol in the original seawater sample was calculated by multiplying the detected HLPC amount (nanomoles) by a concentration factor f_c:

$$f_c = \frac{V_{elute}}{V_{aliquot}} * \frac{V_{vial}}{V_{extract}}$$

Where V_{elute} is the volume eluted from the PPL cartrdige (3 mL), $V_{aliquot}$ the volume of sample added to reaction vessel for cupric oxidation (0.6 mL), V_{vial} is the volume of the HPLC vial (0.200 mL) and $V_{extract}$ is the volume (mL) of original seawater passing through the PPL cartridge.

Total dissolved lignin phenol concentration (nM) was then estimated as the sum of the 11 lignin phenol concentrations (TDLP11). The ratio between the sum of syringyl (S; SAD, SAL, SON) and vanillyl (V; VAD, VAL, VON) phenols (S/V), between cinnamyl (C; CAD, FAD) and V phenols (C/V) and between p-hydroxyl phenols (P; PAD, PAL and PON) and V phenols (P/V) were calculated and used as diagnostic tool for understanding the source and fate of lignin phenol molecules in the ocean (Hernes & Benner, 2003; Kaiser et al., 2017; Lobbes et al., 2000; Mann et al., 2016).

Based on comparisons of salinity, fluorescence intensities and lignin phenol concentrations, five of the lignin samples were removed from the study. The concentration for multiple lignin phenols in these samples were either much higher or lower than other samples with similar salinity or fluorescence intensity for peak C and these samples did therefore not follow the expected depth profile.

N-PLS modelling

Partial Least Square performs simultaneous decomposition of the data in X and Y and creates principal components, named latent variables (LVs), that maximises the covariance between the two datasets (Höskuldsson, 1988). For this study, N-PLS was chosen over traditional PLS regression, since X is a cube consisting of EEMs (three-dimensional and not full rank) with dimensions I x J x K (where I refers to samples, J to emission wavelengths and K to excitation wavelengths) and Y is a matrix with dimensions I x M (where M refers to the number of lignin phenols). X and Y are simultaneously fitted onto each other by use of an iterative algorithm (Höskuldsson, 1988; Wold, 1975). For each component, the LVs in X and Y are rotated so that the scores in Y have the highest correlation (r^2) with the scores in X. The obtained model for a component is then subtracted from X and Y and a new set of LVs is found from the residual data not explained by the previous component(s).

All lignin samples (n=48) were included in the development of the N-PLS model. The data in the X and Y was mean-centered and the samples were split into two sets, i.e., a calibration set (n=24) and a test set (n=24) for each block (X and Y). N-way Partial Least Square models with 1-10 components were built using the calibration set and then applied to the test set for validation. The N-PLS modelling was performed in MatLab (2021a) using the PLS toolbox (PLS_Toolbox 8.6.1 (2018). Eigenvector Research, Inc., 284 Manson, WA USA 98831; software available at http://www.eigenvector.com).

To evaluate the N-PLS models and find the appropriate number of LVs, the explained variance of the calibration set and the Root Mean Squared Error of Prediction (RMSEP) of the test set were used. To evaluate the predictive power of the best N-PLS model for each of the lignin phenols, the predicted values were correlated to their measured values across all samples. To assess the correlation between predicted and measured values for all samples, r^2 and the Relative Root Mean Square Error (RRMSE) were calculated based on the correlation. A more

detailed description of the statistical parameters and how they are calculated is given in the Supplementary Information.

The best N-PLS model was then applied to predict TDLP11 concentrations and predetermined lignin phenol ratios for all 327 samples collected for FDOM. The 327 EEMs were first mean-centred prior to applying the N-PLS model. Samples where the prediction yielded a sum of squares (SSQ) three times above the median SSQ value across all samples were excluded (n=17). Fifteen of the excluded samples had an abnormal high fluorescence intensity with a peak max around Ex=265 and Em=320 (which could be a protein-like fluorescence according to Coble (1996))

Characterisation of hydrographic conditions

In the Fram Strait, there is a clear frontal region at the meridian (0°) separating the two current regimes EGC and WSC (see Figure 1A). First, temperatures for ECG are generally sub-zero, whereas it stays above zero for WSC (see Figure 1C). Higher DOM fluorescence (peak C, excitation: 320-360 and emission: 420-460 according to Coble (2007)) is associated with the EGC, whereas the lowest DOM fluorescence is found in the warmer and saline waters of the WSC (see Figure 1D). However, low DOM fluorescence is also found for depths below 150 m on the East Greenland shelf. WSC consists of one warm water mass (>2° C) with high salinities (>34 ‰), from surface to bottom, namely the AW. In contrast, the ECG can be divided into three water masses (PSW, RAW and MAW) primarily based on salinity (see black lines in Figure 1B for water mass division). The PSW is relatively fresh (salinity <34 ‰) and has sub-zero temperatures (Amon et al., 2003) and is found between the surface and 150 m depth on the East Greenland shelf and slope, however the depth of the layer depends on longitude (between 0-17 °W; see Figure 1B & C). The MAW, running below PSW on the East Greenland shelf and slope, have higher salinities (>34 ‰), but still sub-zero temperatures (Amon et al., 2003). RAW originates directly from AW, as the WSC makes a U-turn in the Fram Strait and is transported southwards with the EGC, whereas the MAW also originates from AW, but has circulated the Arctic Ocean at intermediate depth for about 35 years before returning to the Fram Strait (Wefing et al., 2021). However since RAW and MAW are often hard to distinguish, we will refer to both as MAW in the following text. The deep ridge between 15-17° W is known as the Norske Trough.

Division of samples into water masses and statistical analysis

The measured and predicted samples were divided into groups depending on the water mass they were sampled from (AW, PSW, MAW) based on their hydrographic conditions (see Table 1 for specific numbers). Exclusively for the predicted lignin phenol parameters, a sample group for the Norske Trough surface waters (NTSW, 0-15 m, 15-17° W) was additionally created, since the lignin phenol ratio revealed some kind of uniqueness for this area of the PSW. The TDLP11 and the diagenetic parameters between the different sample groups, hence water masses, were then compared by applying analysis of variance (ANOVA) one-way test to calculate statistical differences between them. All the p-values (p) from the ANOVA one-way test can be seen in Supplementary Information (see Table S2)

Table 1: Division of measured and predicted samples into groups depending on the water mass they belong to. The water masses are defined by characteristics from Amon et al. (2003). AW=Atlantic water, PSW=Polar surface water, MAW=Modified Atlantic water, NTSW=Norske Trough surface water (own definition).

Measured samples					
Water mass	**Salinity limit (‰)**	**Depth limit (m)**	**Westernmost longitude**	**Easternmost longitude**	**No. samples**
AW	>34	>600	0°	10° E	7
PSW	<34	<150	18° W	0	33
MAW	>34	>100	18° W	7° W	3
Predicted samples					
Water mass	**Salinity limit (‰)**	**Depth limit (m)**	**Westernmost longitude**	**Easternmost longitude**	**No. samples**
AW	>34	>2692	0°	10° E	92
PSW	<34	<150	15° W	0°	103
MAW	>34	>100	18° W	15° W	16
NTSW	<34	<15	18° W	15° W	12

RESULTS

Measured TDLP11 and lignin phenol ratios

Measured TDLP11 concentrations in the Fram Strait varied between 0.8 and 4.8 nM (see Figure 2). TDLP11 was lowest for AW in WSC (7.5 °E in Figure 2) where concentrations between 1 and 2 nM were found along the water profile, with a tendency to increase with depth. Associated with the highest DOM, the PSW above the East Greenland slope (6 and 8 °W in Figure 2) was found to have the highest measured concentration of lignin phenols, especially between 0-50 m depth. It was clear from the TDLP11 profiles, that the concentration for the PSW on the East Greenland slope were approximately twice compared to the rest of the PSW

on the East Greenland shelf (longitudes 10-17 °W in Figure 2) and approximately four times higher than AW, and also higher than MAW running below PSW (<150 m). The TDLP11 for PSW was therefore found to be significantly higher than both AW and MAW ($p<0.001$).

The estimated lignin phenol ratios, S/V, C/V and P/V for the measured samples ranged between 0.21-0.83, 0.08-0.35, and 0.82-2.07 respectively. Highest measured ratios for S/V, C/V and P/V (> 0.5, >0.2, and >1.4 respectively) across the Fram Strait were associated with the top 100 m of AW (7.5 °E in Figure 2). However, in Norske Trough at 17 °W the seawater between 25-250 m had comparable P/V ratios (>1.4) to AW surface waters. In AW, the S/V ratio was highest (0.83) in the surface and decreased systematically with depth, while P/V was highest at 25 m (1.76) and first started decreasing after 55 m increased. The C/V ratio in AW, first increased towards a maximum (0.35) at 25 m, but afterwards decreased with depth below that.

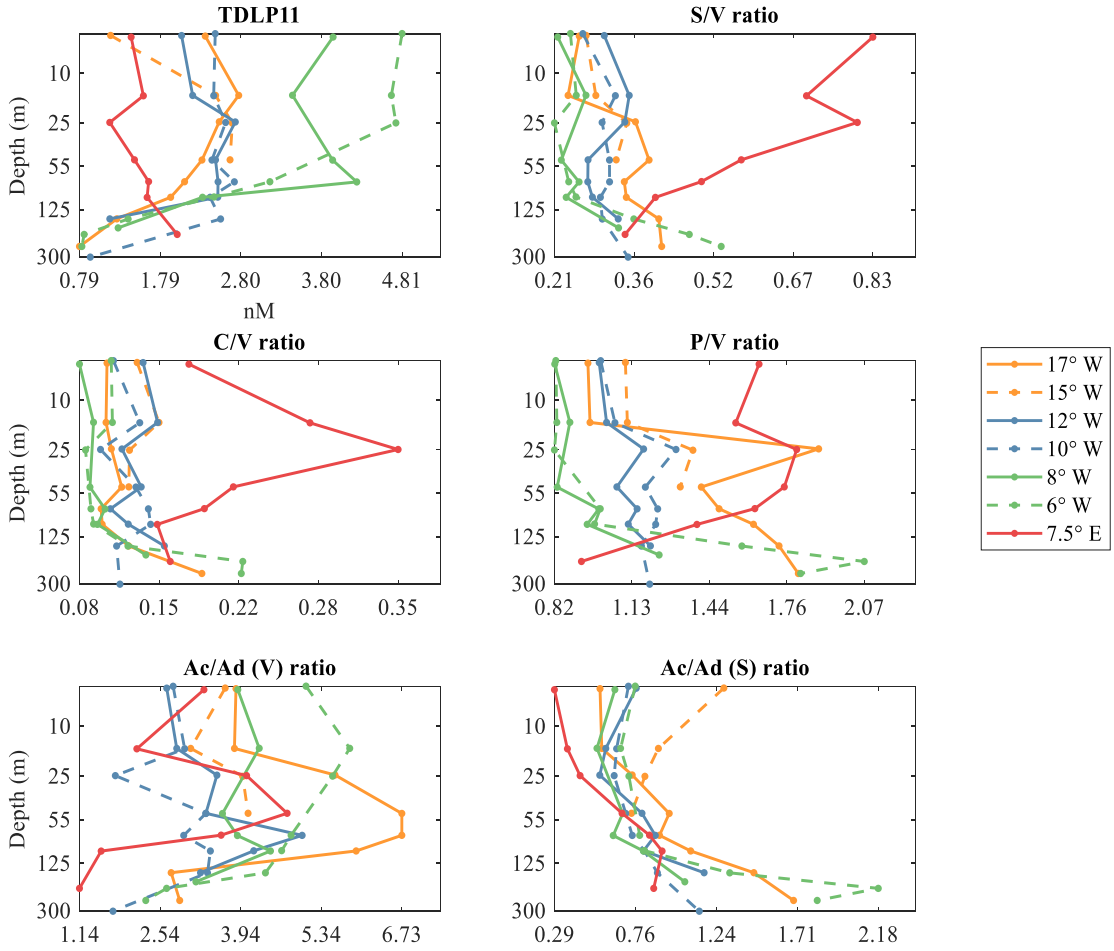


Figure 2: Profiles of measured lignin phenol parameters across the Fram Strait. TDLP11 is the total lignin concentration (sum of all eleven lignin phenols.) and the diagenetic ratios are used to investigate source and diagenesis of lignin. S=syringyl phenols, V=vanilyl phenols, C=cinnamyl phenols, P=p-hydroxy phenols,

The S/V, P/V and C/V ratios in the MAW and PSW (see 6-17° W in Figure 2) were significantly lower ($p<0.001$ for all ratios) than AW and more uniform along the water profile. The uniformity of the lignin ratio profiles, on the East Greenland shelf and slope, was reflected by the fact that the ratios for MAW and PSW was not significant different ($p>0.16$ for all ratios).

Below a 100 m, the S/V and C/V ratios across the transect seemed to be comparable. However, statistical test showed that the S/V and C/V ratio for MAW and AW actually were significant different ($p<0.001$ for both ratios). In contrast, the P/V ratio were not similar at depth across the transect, as the P/V profile for Norske Through (see 17° W in Figure 2) begins to increase and deviate from the other longitudes already at 25 m depth. Additionally, did the P/V profile on the East Greenland slope at 6° W also increase drastically from 100 m depth.

The Ad/Al (V) ratios estimated from measured lignin phenols ranged from 1.1 to 6.7 and did not show a clear trend with longitude nor depth (see Figure 2). The Ad/Al (V) ratios seemed to varied randomly with depth and for AW (7.5 °E) and Norske Trough (17° W) there seemed to be a sudden increase in Ad/Al (V) around 50 m depth. However, PSW was still found to hold significant different Ad/Al (V) ratios from MAW ($p<0.001$), but not compared to AW ($p=0.61$). Nonetheless, the Ad/Al (V) ratios for MAW and AW were surprisingly found to significantly different ($p=0.01$). In contrast to the lack of trend for Ad/Al (V), the Ad/Al (S) ratio for almost all stations across the transect increased with depth in a similar fashion (see Figure 2). However, for station 15° W the Ad/Al (S) ratio decreased with depth, rather than increase. None of the water masses differed significantly from each other when comparing the Ad/Al (S) values ($p>0.07$).

Mixing between water masses

To investigate the mixing between water masses in the Fram Strait, S/V ratios were plotted against C/V ratios (Figure 3A & B). Reference lines are drawn between the 25 m samples from 7.5° E, 6° W , 17° W and the 15 m sample from 15° W, which represent the most extreme lignin phenol ratios for the measured samples. The measured sample with the maximum ratios for both S/V and C/V represents surface water from AW in WSC (7.5° E, 25 m) and the sample with the minimum ratios originates from PSW in the EGC, more particular on the East Greenland slope (6° W, 25 m). From the provenance plot (Figure 3A) it can been seen that

most of seawater samples taken across the EGC (6-17° W) are placed in the lower reference triangle and clearly deviate in values from the samples taken for AW across all depths (7.5 ° E). However, several samples from EGC that places in the upper reference triangle, which seems to be the ones taken from MAW below PSW (>150 m depth, see Figure 3B). The AW samples taken from WSC cluster less than those in EGC and the ratios decrease with depth from 25 m and down to 250 m, slowly entering the upper reference triangle. The surface samples from AW deviate from the other ones in upper reference triangle of the provenance plot.

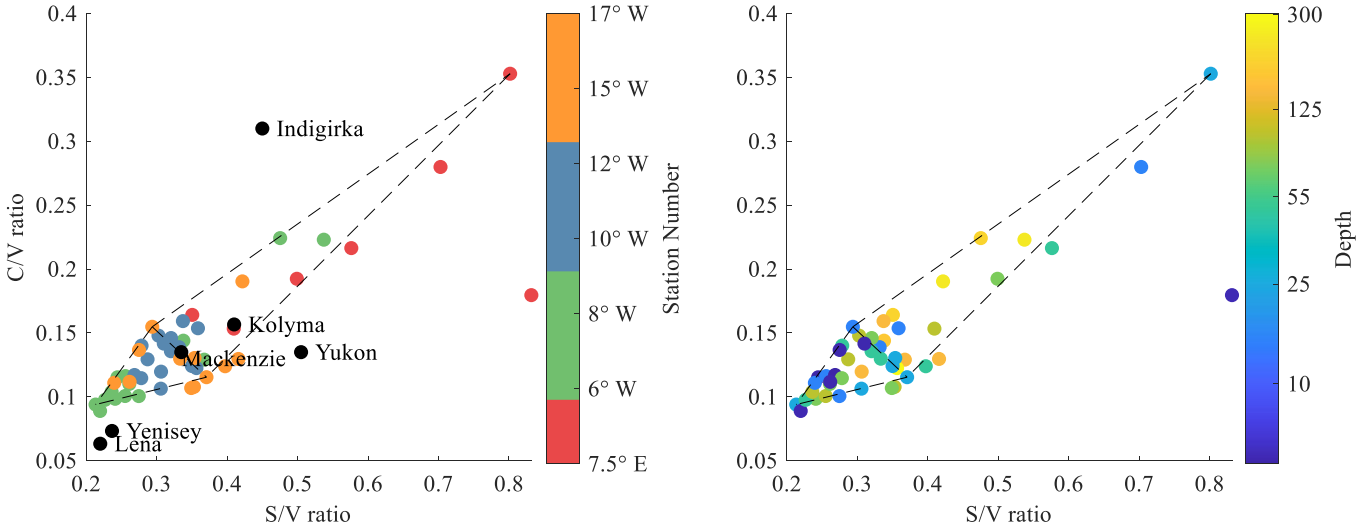


Figure 3: Mixing plots between endmember ratios in the Fram Strait, based on C/V and S/V ratios. The points are either colored by their station longitude (A) or by their depth (B). The black dots in (A) are river endmember ratios based on the average value from other studies (Amon et al., 2012; Lobbes et al., 2000; Mann et al., 2016). The dashed lines are reference lines between the extremes, 25 m samples from 7.5° E, 6° W, 17° W and the 15 m sample from 15° W.

Fluorescence properties of contrasting endmember waters

As with TDLP11, the DOM fluorescence intensity and characteristics changes across the Fram Strait (Figure 1D, and Figure 4). At higher salinities in AW, the DOM fluorescence intensity is much lower and the fluorescence signal is shifted to lower excitation and emission wavelengths. There is a peak centred around excitation 275 nm and emission 320 nm in AW and MAW at 10° W, which is not prevalent in the PSW and MAW at 17° W. As seen from Figure 1D, DOM fluorescence in the PSW vary in intensity but is more or less similar in spectral character (see PSW EEMs in Figure 4). The EEMs for PSW, and also to some extent

for MAW at 17° W, have a peak centred around excitation 305 nm and emission 405 nm, which is not apparent in AW and MAW.

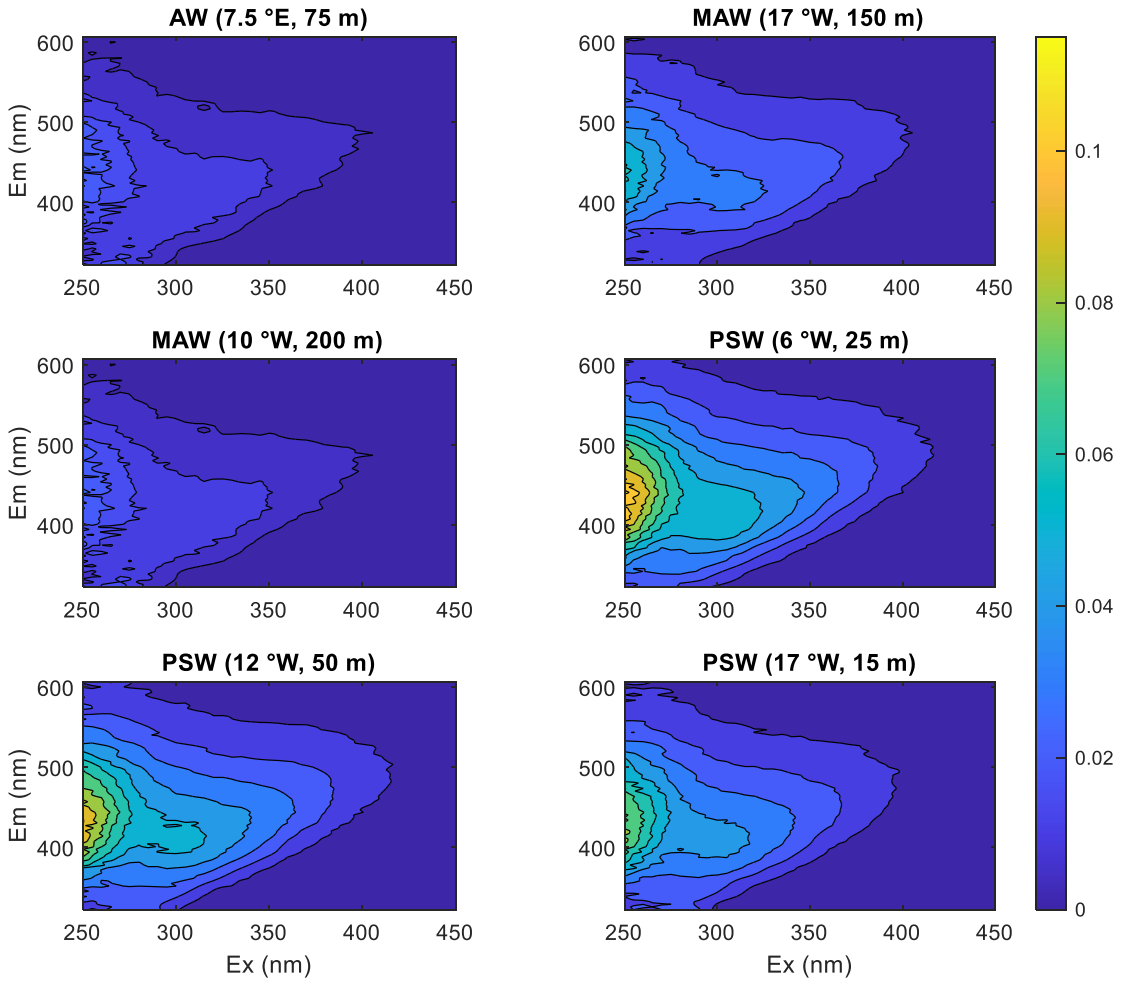


Figure 4: Representative examples of EEMs from contrasting locations along the transect. Fluorescence is in Raman Units (R.U.). Note the different intensity scale between samples.

Linking fluorescence to lignin phenol content

A linear regression analysis between the DOM fluorescence and lignin phenol parameters was carried out to assess if the quantitative and qualitative changes in fluorescence could be linked to differences in lignin content. The highest r^2 for any excitation and emission pair in the EEMs was found to vary between 0.40 and 0.66 depending on lignin parameter (see Figure S1). For most of the lignin parameters, broad regions across the EEMs were equally correlated across an area covering excitation 250-380 nm and emission 340-575 nm and no particular wavelength region (peak) stood out. However, there was a tendency for higher r^2 values toward lower excitation and emission wavelengths.

The appropriate number of LVs for predicting lignin phenols based on fluorescence using N-PLS was determined by considering the variance explained and RMSEP (see Figure 5). Using six LVs explained 99% and 88% of the variance in X and Y, respectively, and gave the third lowest sum of RMSEP across all the lignin phenols. From the Hotelling plot (see Figure S3), one sample in calibration and another sample in the validation dataset were found to be outliers and subsequently removed from the analysis.

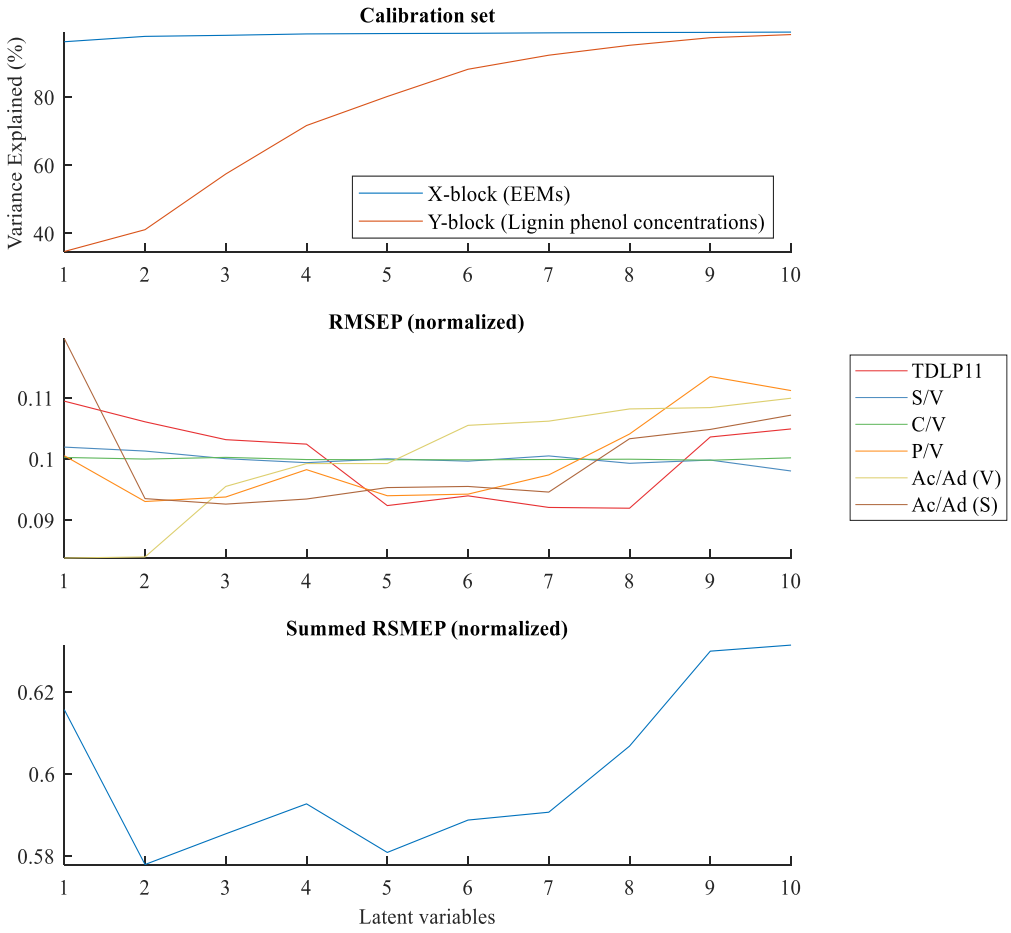


Figure 5: Statistics for application of the N-PLS model to describe to data in the calibration set (A) and afterwards predict the data in test set (B,C) with increasing amount of latent variables.

Predicting the concentration of TDPL11 and lignin phenol ratios using the best N-PLS model (with six LVs) on the EEMs in both the calibration and test set, and then correlating the predicted values to the measured ones, yielded r^2 values between 0.24 and 0.76 depending on the parameter (see Figure 6). The RRMSE between predicted values and the measured ones across the calibration and test set ranged between 15-36% (see Figure 6). For TDLP11 the r^2

and a RRMSE were found to be 0.76 and 20% respectively, whereas S/V, C/V, P/V gave r^2 and RRMSE values of 0.67, 0.54, 0.68 and 23, 24, 15% respectively. Ad/Al (V) and Ad/Al (S) yielded the lowest values with r^2 of 0.24 and 0.24, respectively, and RRMSE of 36% and 32%, respectively. Nevertheless, the regression between the predicted and measured ratios of Ad/Al (V) and Ad/Al (S) were not found to be significant (p-values > 0.52). For TDLP11, S/V, C/V and P/V the regression line was close the 1:1 line.

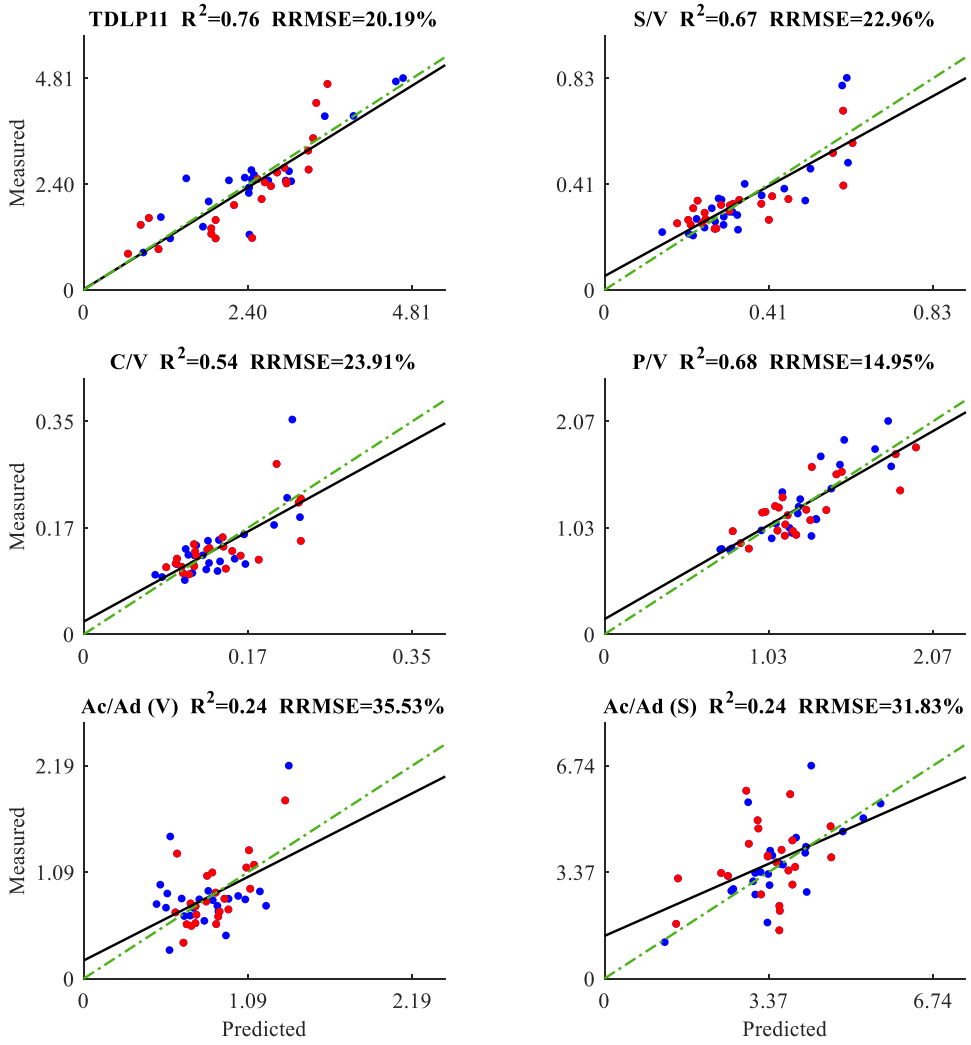


Figure 6: Measured values plotted against predicted values for the calibration (blue points) and the test set (red points) together, including the regression line (black) and a diagonal line (green).

The RRMSE indicates the margin of error using the regression from the N-PLS models to predict TDLP11 and lignin phenols ratios extrapolated from EEMs. Samples within the close-

to-zero range of the regression line can therefore potentially lead to prediction of negative values.

To challenge the correlation power of the selected N-PLS for prediction of TDLP11 and lignin phenol ratios, the calibration and test set were swapped around, so that the test set was now the calibration set and vice versa. The new N-PLS model, built on the swapped sets, gave r^2 and RRMSE values very similar to the original N-PLS model for all parameters except the P/V, which had higher r^2 and lower RRMSE values (see Split half 2 in Table S1). In addition, cross-validation method leave-one-out (LOO) was also performed on the data to produce a N-PLS model. The LOO showed to perform similar to the original N-PLS model for TDLP11, S/V, Ad/Al (V), however had better fit for C/V and Ad/Al (S), but worse for P/V (see LOO Table S1).

Predicted TDLP11 and lignin phenol ratios

First, fourteen EEMs could not be adequately modelled (excessively high SSQ), having fluorescence properties that deviated too much from the calibration set, and were therefore excluded. For the remaining samples, the N-PLS model overall performed very well, placing most samples (n=313) within the boundaries of the Hotelling plot (Figure S3).

With the much larger spatial coverage of sampling from the FDOM measurements, predicted TDLP11 concentrations and lignin phenol ratios could be plotted as section plots (see Figure 7) and gave a higher resolution, as every longitude degree is included, extending the information of AW in WSC from only one station (7.5° E, measured data) to 11 stations from the middle of the Fram Strait to the coast of Svalbard (0-10° E). For the water masses in ECG we also increase the spatial resolution with 14 stations, including every degree longitude from the middle of Fram Strait and across the entire East Greenland shelf (1-17° W).

The predictions of TDLP11 shows that significantly higher concentrations were predicted for PSW across the entire ECG (1-17° W) compared to predicted concentrations across MAW ($p<0.001$) and across AW in WSC ($p<0.001$). TDLP11 concentrations for MAW and AW did however not significantly differ from each other ($p=0.18$). PSW on the East Greenland slope (4-9° W) yielded the highest predicted concentrations across the whole transect. The TDPL11 concentrations within the AW were very uniform across longitudes and depths, but seemed to increase in the surface waters (0-200 m depth) eastwards from 7.5 °E to 10 °E. The large-scale

pattern in predicted TDLP11 concentrations resembles that for the intensity of peak C fluorescence (Figure 1D)

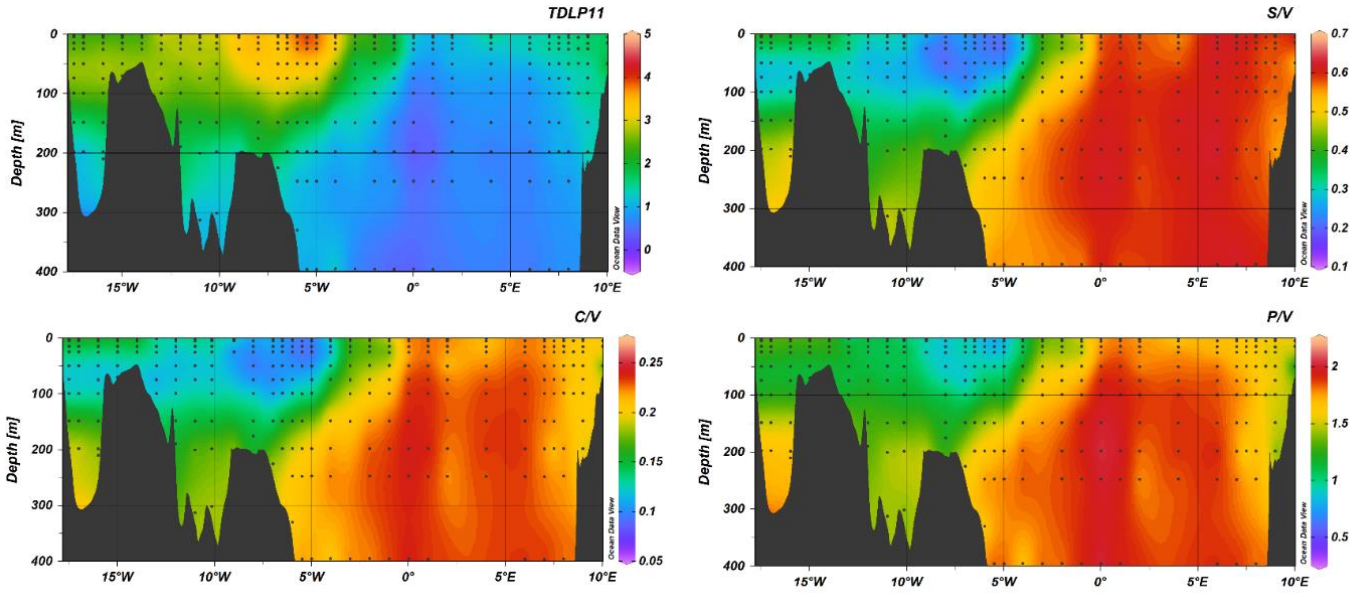


Figure 7: Predicted TDLP11 concentration (nM) and C/V, S/V and P/V ratios across the Fram Strait at 0-400 m depth

Predicted S/V, C/V, P/V ratios could also be plotted as section plots (see Figure 7). The predicted S/V and C/V ratios were highest for the PSW and lowest for AW whereas P/V ratio had the opposite trend with highest values for AW. As with TDLP11, AW holds very uniform S/V (0.6-0.7) and C/V (0.2-0.25) ratios. However, in AW there seemed to be slightly lower C/V ratios close to Svalbard (8-10 °E). For EGC, the S/V and C/V varied more with both depth and longitude and had the lowest ratios in the core of the PSW on top of the East Greenland slope (4-9° W) and highest ratios below in MAW. Finally, the predicted S/V, C/V and P/V ratios for PSW showed be significantly different from AW and MAW ($p<0.001$). The predicted S/V and C/V ratios for AW and MAW also differed significantly for each other ($p<0.001$), but not for the P/V ratio ($p=0.52$). Interestingly, the N-PLS predictions revealed that the NTSW above Norske Trough (17° W) holds significantly different S/V, C/V and P/V ratios ($p<0.05$) compared to the rest of the PSW.

Since the regression between the measured and predicted ratios for Ad/Al (V) and Ad/Al (S) was found not to be significant, prediction using these relationships were not performed across the Fram Strait.

DISCUSSION

Fluorescence characterization of water masses

The water masses in Fram Strait can be divided based on the intensity and character of FDOM (Figure 1D, and Figure 4). The PSW water masses yield the highest fluorescence intensities of peak C, whereas AW the lowest, and MAW an intermediate fluorescence. MAW at 10° W contained very comparable fluorescence intensity to AW reflecting its origins.. The MAW at 17° W was more similar to PSW and may in fact have circulated the Arctic Ocean for a long period. The large difference in FDOM intensity and character between PSW and AW is probably due to terrestrial DOM from Arctic rivers (Amon et al., 2003; Kaiser et al., 2017; Stedmon et al., 2011). Amon et al. (2003) found that the position of peaks in EEMs for PSW in the Fram Strait was very similar to the ones found from three Arctic rivers (Ob, Yenisei and Mackenzie) and Gonçalves-Araujo et al. (2016) found that specifically the fluorescence of two humic-like components were associated with PSW in the Fram Strait. For AW, Granskog et al. (2012) found a higher spectral slope in the 275-295 nm region of the absorbance spectrum compared to PSW, which indicate more photochemically changed material and a different CDOM source, which may explain why the EEMs for AW and MAW at 10° W are slightly shifted towards lower excitation and emission wavelengths. MAW has a higher fluorescence intensity and slightly changed character compared to RAW, which also reflects its transit route through the Arctic Ocean (Figure 1D, and Figure 4).

Predicting lignin phenol content from fluorescence

Previous studies (Hernes et al 2009; Fichot et al 2016) have been able to predict TDLP from spectral measurements at very high accuracy yielding r^2>0.90, which is higher than what was found in this study for our N-PLS model. However, these studies used riverine and coastal samples (3-15 times higher lignin concentration and a magnitude higher fluorescence intensity for the riverine samples) to establish correlation between lignin phenols and spectral data. Especially in riverine environments, it is more likely that fluorescence EEMs carry measurable lignin-specific spectral features that N-PLS models can build on. However, Fichot et al. (2016) found a model that could not predict lignin in seawater adequately when the TDLP concentration was less than 6 nM, which complicates its use for open ocean samples. Our model therefore represents an extension of these efforts into the challenging oceanic realm with all TDLP11 concentrations below 6 nM. When concentrations reach oceanic levels,

fluorescence-based predictions become more challenging and model performance may suffer. Nonetheless, correlation between predicted and measured S/V and C/V still showed to be comparable the ones found by Hernes et al. (2009), 0.74 and 0.50 respectively.

The major challenge for our N-PLS model is likely the AW samples. The EEMs taken from AW have considerably lower signal-to-noise ratios, making it difficult for any model to achieve robust predictions without significant bias from measurement noise. Another explanation can be poor coverage of the diversity of the lignin-derived material across the Fram Strait. The calibration dataset mainly consisted of samples from the Greenland shelf waters and could be biased towards this tDOM and FDOM. This could be examined by expanding the data coverage of the lignin phenol measurements, to then provide more confidence in the model predictions.

Lignin phenol distribution across the Fram Strait

Comparing the general trends seen for the predicted parameters in the section plot (see Figure 7) with the trends in the measured profiles (see Figure 2) revealed that the predicted data mirror the trends very well for most of the transect. However, for the AW the increase in TDLP11 concentrations seen at deeper waters from the measured profile (Figure 2) is not reflected in the section plot (see Figure 7). From the TDLP11 section plot, it additionally seems from the concentration follows the fluorescence intensity of peak C.

Both the results from direct measurement of lignin phenols and N-PLS predicted lignin phenols showed larger TDPL11 concentrations in the Arctic outflow than the Atlantic inflow. The measurements agree with earlier findings. Kaiser & Benner (2012) reported 1.9 nM for TDLP11 in Arctic Ocean surface waters and Benner et al. (2005) found approximately 2 nM (for S and V phenols) in the EGC south of the Fram Strait. The low TDLP11 concentrations observed within the AW is consistent with previous report from the Atlantic Ocean (TDLP6 < 1 nM) (Opsahl et al., 1999; Opsahl & Benner, 1997). Kaiser et al. (2017) found TDLP9 (sun of P, V and S phenols) to vary between 1.4 to 16.5 nM for shelf break waters of the East Siberian Sea in waters above 150 m, where highest concentration were found for samples closest to the shelves. This indicate that the TDLP11 concentrations found in this study for the core of PSW in the Fram Strait fits well with the TDLP9 concentrations leaving the East Siberian Sea and underline the connectivity between the two regions. Additionally, did Kaiser et al. (2017) find 1.2 and 2.6 nM for waters below 150 m, which fits with the values for MAW found here in our study.

The N-PLS predictions revealed that MAW only has a minor contribution of terrestrial DOM, even though it has travelled the Arctic Ocean over a decade, as the TDPL11 concentrations are similar to the inflowing AW. From salinity and the fluorescence intensity of peak C (Figure 1B & 1D) it is also clear that the MAW has not mixed with fresh water from the Arctic rivers in the same extent as PSW above and this lack of mixing therefore explain the lower concentrations of lignin phenols. However, the MAW was found to be more similar S/V and C/V ratios to PSW than AW and MAW must therefore have picked up some lignin from Arctic rivers along the circulation.

Source and fate of lignin across Fram Strait

S/V and C/V ratios have traditionally been used for determination of source material (Amon et al., 2012; Hedges & Mann, 1979) whereas the P/V ratio also indicates photodegradation (Benner & Kaiser, 2011; Kaiser et al., 2017). However, the S/V ratio has also be shown to be biased by photodegradation (Hernes & Benner, 2003; Opsahl & Benner, 1998; Spencer et al., 2009), which can limit its use for determining lignin sources in water masses that have been exposed to long-term sunlight. For the Arctic Ocean, exposure to sunlight may not be an important factor, since ice coverage prevents much of the region from ultraviolet radiation. Nevertheless, in general all lignin phenols are to a certain degree susceptible to light, where the decay rates for the phenolic subunits are found to be in the order S > V > P (Benner & Kaiser, 2011), whereas C phenols have been found to be the most resistant to photodegradation (Spencer et al., 2009).

For AW, measured and predicted S/V, C/V and P/V ratios were among the highest values across the Fram Strait. S/V ratios for the Pacific and the Atlantic Ocean are expected to range between 0.1 and 0.4 (Kaiser et al., 2017; Opsahl & Benner, 1997) and these ratios are therefore lower than those found in this study. However, the measured S/V ratios for AW at depth <100 m tend to be more similar to other findings. The high S/V ratios found in our study, could indicate that particularly the surface waters (0-25 m) for AW transport some lignin material, most likely from angiosperm sources (Hedges & Mann, 1979) before reaching the Fram Strait. The predictions revealed an elevated terrestrial signal in the surface of AW closest to the coast of Svalbard, despite very low peak C fluorescence intensities for these water masses which did not stand out from the rest of the AW (see Figure 1D). The source of this lignin could be derived from further south, along the Norwegian coastal current, with a terrestrial DOM contribution

from Norwegian rivers and possibly, from the Baltic outflow, which has similar S/V ratios (Osburn & Stedmon, 2011). In general. we found higher P/V ratios for AW compared to PSW, which could indicate that the lignin material in AW is older and more photodegraded. The P/V values for AW found in our study are however slightly lower than the ones found by Kaiser et al. (2017). Nevertheless, similar to Fichot et al. (2016), an inverse relationship between P/V and TDLP11 for the measured samples was observed across the Fram Strait (Figure S2), indicating that lower concentrations of lignin in saline waters could be caused by higher exposure to photodegradation likely due to the absence of sea-ice coverage.

The PSW above the East Greenland slope was found to hold the lowest measured and predicted S/V and C/V ratios, significantly lower compared to AW indicating another source material, which is most likely due to incorporation of terrestrial DOM from Arctic rivers. DOM from Arctic rivers carry a mixture of tDOM from tundra and boreal sources with S/V and C/V ratios between 0.12-0.72 and 0.06-0.86 (Amon et al., 2012; Lobbes et al., 2000; Mann et al., 2016; Opsahl et al., 1999). The Lena and the Yenisei are the two largest contributors of tDOM to the Arctic Ocean with export of 4.7 and 7.3 Tg carbon yr^{-1}, respectively (Opsahl et al., 1999). The S/V ratios reported for Lena and Yenisei are 0.19-0.31 and 0.11-0.28, respectively, whereas the C/V ratios are 0.06-0.07 and 0.08, respectively (Amon et al., 2012; Lobbes et al., 2000; Mann et al., 2016). Based on the measured results, we found that the sample with highest TDLP11 and lowest salinity (25 m at 6° W) had S/V and C/V ratios of 0.21 and 0.08, respectively, and are therefore very similar to the values found for the Lena and Yenisei (see Figure 3). In turn, the predicted and measured P/V ratios (0.80-0.90) for the PSW above the East Greenland slope, are very similar to the ratios found by Kaiser et al. (2017) for the Laptev Sea and East Siberian Sea, which are the major shelf areas where the largest Eurasian rivers (Yenisey, Lena and Kolyma) discharge into. The similarities between the lignin phenol ratios indicate the connectivity between the East Siberian shelf DOM and terrestrial DOM found in ECG in the Fram Strait and additionally indicate that limited bio- and photodegradation occurs during its transport across the central Arctic Ocean under the sea-ice.

From the predictions of lignin phenol ratios, it seems that the NTSW above Norske Trough appears to derive from a different source compared to the rest of PSW, since they were found to have significant different S/V, C/V and P/V ratios. The predicted S/V and C/V ratios found in NTSW match more with lignin phenols ratios the rivers Kolyma (0.31-0.41 and 0.10-0.21) and Mackenzie (0.33 and 0.10) (Amon et al., 2012; Lobbes et al., 2000). The P/V ratios for Kolyma and Mackenzie are 1.02 and 0.63 respectively (Amon et al. 2012) are however lower

than for NTSW. The lower P/V ratios in NTSW could indicate degradation of lignin phenols on the journey from their source rivers to the Fram Strait.

As with TDLP11, AW holds very uniform S/V and C/V ratios across, underlining that WSC carry one large water mass with the same fingerprint across longitudes and depth.

Water mass differentiation and mixing based on lignin phenols

From the results it was clear that PSW and AW differed from each other. However, it was also clear that MAW was more similar to PSW than AW based on the lignin phenol parameters presented here.

From the predicted ratios in the section plots and from the statistical analysis between NSTW and the rest of PSW, it seemed that two distinct water masses exist in the surface water of EGC carrying different terrestrial signals based on lignin phenol parameters. From the provenance plot (see Figure 3A) the measured ratios in ECG (both PSW and MAW) were constrained in the lower mixing triangle, with some exceptions. While the measured ratios from PSW above the East Greenland slope (green dots in Figure 4A) clearly deviated from the rest of ratios in the ECG, the PSW above the East Greenland shelf and Norske Trough (respectively blue and orange dots in Figure 3A) are more spread out across the lower mixing triangle. The lignin phenol ratio provided by the N-PLS model clearly shows qualitative differences in the high lignin water on the shelf (Figure 7).

Towards in situ applications

For the approach to provide maximum benefit, it should be applicable to future in-situ sensors featuring stronger LEDs and spectral detectors (Zielinski et al., 2018). This would vastly expand the potential data coverage. For this, it is necessary to consider the constraints of sensor design when establishing a model. Sensors typically need one LED per excitation wavelength and have a limited number of such LEDs due to power constraints.

To satisfy such constraints, the EEMs were trimmed to four emission slices (320–620 nm) at excitation wavelengths 255, 275, 315 and 360 nm. These excitation wavelengths were selected since light sources at wavelengths are commercially available. Performing the N-PLS analysis with reduced data gave the best fit and prediction for X and Y using five LVs, resulting in highest variance explained for the calibration set (99% for X and 89% for Y) and lowest

cumulative RMSEP for the test set. The reduction in excitation wavelengths did not have a large impact on the correlation between measured and predicted concentrations and the r^2 values were therefore comparable (within +/- 0.05, Table S3).

The reasonably good prediction opens the possibility of designing lignin-predicting in-situ fluorescence sensors, which can be attached to autonomous platforms such as moorings, profiling floats and gliders and thereby provide measures of lignin distributions at higher resolution, longer time periods and larger spatial scales.

CONCLUSION

TDLP11 concentrations across the Fram Strait vary between 0.8 and 4.8 nM with highest concentrations in the high DOM waters on the shelf originating from the Arctic Ocean. The lignin phenol composition indicate that the source material and the diagenesis of lignin-derived material in PSW and AW differs, and that PSW is similar to Arctic rivers and likely less photodegraded. Based on the measured and the predicted lignin phenol ratios, it appears that the terrestrial material in PSW originates from at least two different endmembers, with Lena and Yenisei dominating the waters with the highest lignin concentration. Additionally, the predicted lignin phenol parameters revealed that the seawater closest to the coast of Svalbard deviates from the rest of the AW in WSC. Due to the scarcity of lignin phenol measurements, the identification of endmembers and mixing processes between them proved challenging. Yet, with the help of predicted lignin phenol concentrations based on DOM fluorescence, resolution and coverage was improved drastically and mixing processes and distinct tDOM sources better resolved. The fluorescence N-PLS approach holds promise for routine monitoring the changing quantity and composition of tDOM exported from the Arctic, allowing for a greater spatial and temporal coverage of data.

Furthermore, it was found that reducing the fluorescence data to only four excitation wavelengths did not reduce the predictive power of the N-PLS model. This can therefore help to pave the way for development and application of in-situ sensors mounted on gliders, buoys and Agro floats, to trace lignin and thereby terrestrial DOM throughout the ocean.

REFERENCES

Amon, R. M. W., Budéus, G., & Meon, B. (2003). Dissolved organic carbon distribution and origin in the Nordic Seas: Exchanges with the Arctic Ocean and the North Atlantic. *Journal of Geophysical Research: Oceans*, *108*(C7), 3221. https://doi.org/10.1029/2002JC001594

Amon, R. M. W., Rinehart, A. J., Duan, S., Louchouarn, P., Prokushkin, A., Guggenberger, G., Bauch, D., Stedmon, C., Raymond, P. A., Holmes, R. M., McClelland, J. W., Peterson, B. J., Walker, S. A., & Zhulidov, A. v. (2012). Dissolved organic matter sources in large Arctic rivers. *Geochimica et Cosmochimica Acta*, *94*, 217–237. https://doi.org/10.1016/j.gca.2012.07.015

Bai, X. M., Liu, T., Liu, D. L., & Wei, Y. J. (2018). Simultaneous determination of α-asarone and β-asarone in Acorus tatarinowii using excitation-emission matrix fluorescence coupled with chemometrics methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *191*, 195–202. https://doi.org/10.1016/J.SAA.2017.10.011

Benner, R., & Kaiser, K. (2011). Biological and photochemical transformations of amino acids and lignin phenols in riverine dissolved organic matter. *Biogeochemistry*, *102*(1), 209–222. https://doi.org/10.1007/S10533-010-9435-4/FIGURES/7

Benner, R., Louchouarn, P., & Amon, R. M. W. (2005). Terrigenous dissolved organic matter in the Arctic Ocean and its transport to surface and deep waters of the North Atlantic. *Global Biogeochemical Cycles*, *19*(2), 1–11. https://doi.org/10.1029/2004GB002398

Bro, R. (1996). MULTIWAY CALIBRATION. MULTILINEAR PLS. *JOURNAL OF CHEMOMETRICS*, *10*, 47–61.

Bro, Rasmus. (1998). *Multi-way analysis in the food industry : models, algorithms, and applications*.

Bruhn, A. D., Stedmon, C. A., Comte, J., Matsuoka, A., Speetjens, N. J., Tanski, G., Vonk, J. E., & Sjöstedt, J. (2021). Terrestrial Dissolved Organic Matter Mobilized From Eroding Permafrost Controls Microbial Community Composition and Growth in Arctic Coastal Zones. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.640580

Coble, P. G. (1996). Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Marine Chemistry*, *51*(4), 325–346. https://doi.org/10.1016/0304-4203(95)00062-3

Coble, P. G. (2007). Marine Optical Biogeochemistry: The Chemistry of Ocean Color. *Chem. Rev.*, *107*, 402–418. https://doi.org/10.1021/cr050350

Cory, R. M., Crump, B. C., Dobkowski, J. A., & Kling, G. W. (2013). Surface exposure to sunlight stimulates CO2 release from permafrost soil carbon in the Arctic. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(9), 3429–3434. https://doi.org/10.1073/PNAS.1214104110

Cory, R. M., Ward, C. P., Crump, B. C., & Kling, G. W. (2014). Sunlight controls water column processing of carbon in arctic fresh waters. *Science*, *345*(6199), 925–928. https://doi.org/10.1126/SCIENCE.1253119

Dittmar, T., Koch, B., Hertkorn, N., & Kattner, G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnology and Oceanography: Methods*, *6*(6), 230–235. https://doi.org/10.4319/LOM.2008.6.230

Fichot, C. G., & Benner, R. (2012). The spectral slope coefficient of chromophoric dissolved organic matter (S 275-295) as a tracer of terrigenous dissolved organic carbon in river-influenced ocean margins . *Limnology and Oceanography*, *57*(5), 1453–1466. https://doi.org/10.4319/lo.2012.57.5.1453

Fichot, C. G., Benner, R., Kaiser, K., Shen, Y., Amon, R. M. W., Ogawa, H., & Lu, C. J. (2016). Predicting dissolved lignin phenol concentrations in the coastal ocean from chromophoric dissolved organic matter (CDOM) absorption coefficients. *Frontiers in Marine Science*, *3*(FEB). https://doi.org/10.3389/fmars.2016.00007

Gonçalves-Araujo, R., Granskog, M. A., Bracher, A., Azetsu-Scott, K., Dodd, P. A., & Stedmon, C. A. (2016). Using fluorescent dissolved organic matter to trace and distinguish the origin of Arctic surface waters. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP33978

Gonçalves-Araujo, R., Stedmon, C. A., de Steur, L., Osburn, C. L., & Granskog, M. A. (2020). A Decade of Annual Arctic DOC Export With Polar Surface Water in the East Greenland Current. *Geophysical Research Letters*, *47*(20). https://doi.org/10.1029/2020GL089686

Granskog, M. A., Pavlov, A. K., Sagan, S., Kowalczuk, P., Raczkowska, A., & Stedmon, C. A. (2015). Effect of sea-ice melt on inherent optical properties and vertical distribution of solar radiant heating in Arctic surface waters. *Journal of Geophysical Research: Oceans*, *120*(10), 7028–7039. https://doi.org/10.1002/2015JC011087

Granskog, M. A., Stedmon, C. A., Dodd, P. A., Amon, R. M. W., Pavlov, A. K., de Steur, L., & Hansen, E. (2012). Characteristics of colored dissolved organic matter (CDOM) in the Arctic outflow in the Fram Strait: Assessing the changes and fate of terrigenous CDOM in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *117*(12). https://doi.org/10.1029/2012JC008075

Green, N. W., McInnis, D., Hertkorn, N., Maurice, P. A., & Perdue, E. M. (2015). Suwannee River natural organic matter: Isolation of the 2R101N reference sample by reverse osmosis. *Environmental Engineering Science*, *32*(1), 38–44. https://doi.org/10.1089/ees.2014.0284

Haine, T. W. N., Curry, B., Gerdes, R., Hansen, E., Karcher, M., Lee, C., Rudels, B., Spreen, G., de Steur, L., Stewart, K. D., & Woodgate, R. (2015). Arctic freshwater export: Status, mechanisms, and prospects. *Global and Planetary Change*, *125*, 13–35. https://doi.org/10.1016/J.GLOPLACHA.2014.11.013

Hedges, J. I., Blanchette, R. A., Weliky, K., & Devol, A. H. (1988). Effects of fungal degradation on the CuO oxidation products of lignin: A controlled laboratory study. *Geochimica et Cosmochimica Acta*, *52*(11), 2717–2726. https://doi.org/10.1016/0016-7037(88)90040-3

Hedges, J. I., & Ertel, J. R. (1982). Characterization of Lignin by Gas Capillary Chromatography of Cupric Oxide Oxidation Products. *Analytical Chemistry*, *54*(2), 174–178. https://doi.org/10.1021/AC00239A007

Hedges, J. I., & Mann, D. C. (1979). The characterization of plant tissues by their lignin oxidation products. *Geochimica et Cosmochimica Acta*, *43*(11), 1803–1807. https://doi.org/10.1016/0016-7037(79)90028-0

Hernes, P. J., & Benner, R. (2002). Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean. In *Deep-Sea Research I* (Vol. 49).

Hernes, P. J., & Benner, R. (2003). Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments. *Journal of Geophysical Research: Oceans*, *108*(9). https://doi.org/10.1029/2002jc001421

Hernes, P. J., Bergamaschi, B. A., Eckard, R. S., & Spencer, R. G. M. (2009). Fluorescence-based proxies for lignin in freshwater dissolved organic matter. *Journal of Geophysical Research: Biogeosciences*, *114*(4). https://doi.org/10.1029/2009JG000938

Hill, V. J. (2008). Impacts of chromophoric dissolved organic material on surface ocean heating in the Chukchi Sea. *Journal of Geophysical Research: Oceans*, *113*(C7), 7024. https://doi.org/10.1029/2007JC004119

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, *2*(3), 211–228. https://doi.org/10.1002/CEM.1180020306

Kaiser, K., & Benner, R. (2012). Characterization of lignin by gas chromatography and mass spectrometry using a simplified CuO oxidation method. *Analytical Chemistry*, *84*(1), 459–464. https://doi.org/10.1021/ac202004r

Kaiser, K., Benner, R., & Amon, R. M. W. (2017). The fate of terrigenous dissolved organic carbon on the Eurasian shelves and export to the North Atlantic. *Journal of Geophysical Research: Oceans*, *122*(1), 4–22. https://doi.org/10.1002/2016JC012380

Kipp, L. E., Charette, M. A., Moore, W. S., Henderson, P. B., & Rigor, I. G. (2018). Increased fluxes of shelf-derived materials to the central arctic ocean. *Science Advances*, *4*(1). https://doi.org/10.1126/SCIADV.AAO1302/SUPPL_FILE/AAO1302_SM_REVISION1.PDF

Kumar, K., & Mishra, A. K. (2012). Quantification of ethanol in ethanol-petrol and biodiesel in biodiesel-diesel blends using fluorescence spectroscopy and multivariate methods. *Journal of Fluorescence*, *22*(1), 339–347. https://doi.org/10.1007/S10895-011-0964-6/TABLES/5

Lawaetz, A. J., & Stedmon, C. A. (2009). Fluorescence intensity calibration using the Raman scatter peak of water. *Applied Spectroscopy*, *63*(8), 936–940. https://doi.org/10.1366/000370209788964548

Lin, Y., Li, X. K., Yan, J. Y., Lei, Y. L., Chen, J., & Li, B. Q. (2022). EEFM combined with N-PLS method for the quantitative determination of psoralen in cosmetics. *Chemical Papers*. https://doi.org/10.1007/S11696-022-02262-5

Lobbes, J. M., Fitznar, H. P., & Kattner, G. (1999). High-performance liquid chromatography of lignin-derived phenols in environmental samples with diode array detection. *Analytical Chemistry*, *71*(15), 3008–3012. https://doi.org/10.1021/ac981175i

Lobbes, J. M., Fitznar, H. P., & Kattner, G. (2000). Biogeochemical characteristics of dissolved and particulate organic matter in Russian rivers entering the Arctic Ocean. *Geochimica et Cosmochimica Acta*, *64*(17), 2973–2983. https://doi.org/10.1016/S0016-7037(00)00409-9

Mann, P. J., Spencer, R. G. M., Hernes, P. J., Six, J., Aiken, G. R., Tank, S. E., McClelland, J. W., Butler, K. D., Dyda, R. Y., & Holmes, R. M. (2016). Pan-arctic trends in terrestrial dissolved organic matter from optical measurements. *Frontiers in Earth Science*, *4*. https://doi.org/10.3389/feart.2016.00025

Matero, S., Poutiainen, S., Leskinen, J., Järvinen, K., Ketolainen, J., Poso, A., & Reinikainen, S. P. (2010). Estimation of granule size distribution for batch fluidized bed granulation process using acoustic emission and N-way PLS. *Journal of Chemometrics*, *24*(7–8), 464–471. https://doi.org/10.1002/CEM.1269

Opsahl, S., & Benner, R. (1995). Early diagenesis of vascular plant tissues: Lignin and cutin decomposition and biogeochemical implications. *Geochimica et Cosmochimica Acta*, *59*(23), 4889–4904. https://doi.org/10.1016/0016-7037(95)00348-7

Opsahl, S., & Benner, R. (1997). Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature 1997 386:6624*, *386*(6624), 480–482. https://doi.org/10.1038/386480a0

Opsahl, S., & Benner, R. (1998). Photochemical reactivity of dissolved lignin in river and ocean waters. *Limnology and Oceanography*, *43*(6), 1297–1304. https://doi.org/10.4319/LO.1998.43.6.1297

Opsahl, S., Benner, R., & Amon, R. M. W. (1999). Major flux of terrigenous dissolved organic matter through the Arctic Ocean. *Limnology and Oceanography*, *44*(8), 2017–2023. https://doi.org/10.4319/LO.1999.44.8.2017

Osburn, C. L., & Stedmon, C. A. (2011). Linking the chemical and optical properties of dissolved organic matter in the Baltic-North Sea transition zone to differentiate three allochthonous inputs. *Marine Chemistry*, *126*(1–4), 281–294. https://doi.org/10.1016/j.marchem.2011.06.007

Pavlov, A. K., Granskog, M. A., Stedmon, C. A., Ivanov, B. v., Hudson, S. R., & Falk-Petersen, S. (2015). Contrasting optical properties of surface waters across the Fram Strait and its potential biological implications. *Journal of Marine Systems*, *143*, 62–72. https://doi.org/10.1016/J.JMARSYS.2014.11.001

Pegau, W. S. (2002). Inherent optical properties of the central Arctic surface waters. *Journal of Geophysical Research: Oceans*, *107*(C10), SHE 16-1. https://doi.org/10.1029/2000JC000382

Sipler, R. E., Kellogg, C. T. E., Connelly, T. L., Roberts, Q. N., Yager, P. L., & Bronk, D. A. (2017). Microbial community response to terrestrially derived dissolved organic matter in the coastal Arctic. *Frontiers in Microbiology*, *8*(JUN). https://doi.org/10.3389/fmicb.2017.01018

Spencer, R. G. M., Aiken, G. R., Wickland, K. P., Striegl, R. G., & Hernes, P. J. (2008). Seasonal and spatial variability in dissolved organic matter quantity and composition from the Yukon River basin, Alaska. *Global Biogeochemical Cycles*, *22*(4). https://doi.org/10.1029/2008GB003231

Spencer, R. G. M., Stubbins, A., Hernes, P. J., Baker, A., Mopper, K., Aufdenkampe, A. K., Dyda, R. Y., Mwamba, V. L., Mangangu, A. M., Wabakanghanzi, J. N., & Six, J. (2009). Photochemical degradation of dissolved organic matter and dissolved lignin phenols from the Congo River. *Journal of Geophysical Research: Biogeosciences*, *114*(G3), 3010. https://doi.org/10.1029/2009JG000968

Stedmon, C. A., Amon, R. M. W., Bauch, D., Bracher, A., Gonçalves-Araujo, R., Hoppmann, M., Krishfield, R., Laney, S., Rabe, B., Reader, H., & Granskog, M. A. (2021). Insights Into Water Mass Origins in the Central Arctic Ocean From In-Situ Dissolved Organic Matter Fluorescence. *Journal of Geophysical Research: Oceans*, *126*(7), e2021JC017407. https://doi.org/10.1029/2021JC017407

Stedmon, C. A., Amon, R. M. W., Rinehart, A. J., & Walker, S. A. (2011). The supply and characteristics of colored dissolved organic matter (CDOM) in the Arctic Ocean: Pan Arctic trends and differences. *Marine Chemistry*, *124*(1–4), 108–118. https://doi.org/10.1016/j.marchem.2010.12.007

Thingstad, T. F., Bellerby, R. G. J., Bratbak, G., Børsheim, K. Y., Egge, J. K., Heldal, M., Larsen, A., Neill, C., Nejstgaard, J., Norland, S., Sandaa, R. A., Skjoldal, E. F., Tanaka, T., Thyrhaug, R., & Töpper, B. (2008). Counterintuitive carbon-to-nutrient coupling in an Arctic pelagic ecosystem. *Nature 2008 455:7211*, *455*(7211), 387–390. https://doi.org/10.1038/NATURE07235

Wefing, A.-M., Casacuberta, N., Christl, M., Gruber, N., Smith, J. N., Wefing, A.-M., Casacuberta, N., Christl, M., Gruber, N., & Smith, J. N. (2021). Circulation timescales of Atlantic Water in the Arctic Ocean determined from anthropogenic radionuclides. *OcSci*, *17*(1), 111–129. https://doi.org/10.5194/OS-17-111-2021

Wold, H. (1975). Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. *Journal of Applied Probability*, *12*(S1), 117–142. https://doi.org/10.1017/S0021900200047604

Yamashita, Y., Fichot, C. G., Shen, Y., Jaffé, R., & Benner, R. (2015). Linkages among fluorescent dissolved organic matter, dissolved amino acids and lignin-derived phenols in a river-influenced ocean margin. *Frontiers in Marine Science*, *2*(NOV). https://doi.org/10.3389/fmars.2015.00092

Yan, G., & Kaiser, K. (2018). Ultralow Sample Volume Cupric Sulfate Oxidation Method for the Analysis of Dissolved Lignin. *Analytical Chemistry*, *90*(15), 9289–9295. https://doi.org/10.1021/acs.analchem.8b01867

Zielinski, O., Rüssmeier, N., Ferdinand, O. D., Miranda, M. L., & Wollschläger, J. (2018). Assessing Fluorescent Organic Matter in Natural Waters: Towards In Situ Excitation–Emission Matrix Spectroscopy. *Applied Sciences 2018, Vol. 8, Page 2685*, *8*(12), 2685. https://doi.org/10.3390/APP8122685

ACKNOWLEDGEMENT

This study was financed by the Independent Research Fund Denmark Grant No. 9040-00266B awarded to C.A.S and the European Union H2020, as part of the EU Project ARICE grant agreement n° 730965, under the ARICE-NoTAC collaboration led by R.G.A. We thank Norwegian Polar Institute (NPI), without whom the cruises to the Fram Strait would not be possible. We would also like to thank the entire research team and the crew aboard R/V Kronprins Haakon during the 2021 expedition for helping out with the scientific and technical aspect of the sampling.

Supplementary Information:
Dissolved lignin phenols across the Fram Strait: Towards in-situ measurements from fluorescence.

Anders Dalhoff Bruhn[1], Urban Wünsch[1], Christopher L. Osburn[2], Rafael Gonçalves-Araujo[1], Mats A. Granskog[3], Colin A. Stedmon[1].

[1]National Institute of Aquatic Resources, Technical University of Denmark, Denmark.

[2] Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, USA.

[3] Norwegian Polar Institute, Fram Centre, Tromsø, Norway.

Correspondence to: adbj@aqua.dtu.dk

Keywords: Lignin phenols, fluorescence, N-PLS, Fram Strait, terrestrial DOM, PLS

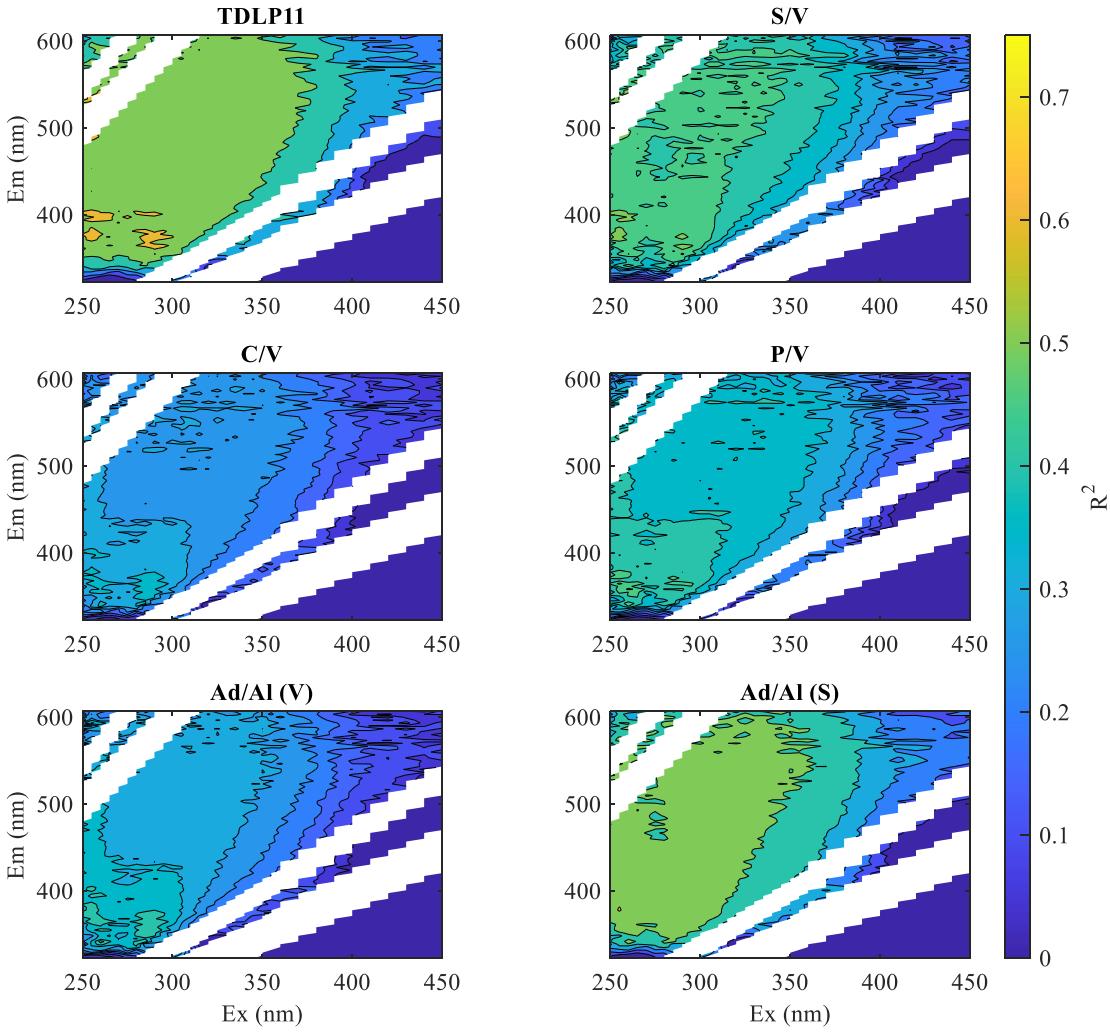


Figure S1: Linear correlation (r^2) between each emission and excitation pair across EEMs and measured lignin phenols parameters (n=46).

Table S1: N-PLS models were created and validated in three different constellations. For split half 1, the 46 measured lignin samples were split into two sets, one calibration and one test. For split half 2, the two sets were swapped around before running N-PLS. In addition to splitting the data in two sets, the leave-one-out (LOO) cross validation was applied. The r^2 and RRMSE values are estimated from performing linear regression between measured and predicted values. NB! Split half 1 is the model used in the article for prediction of lignin parameters across the Fram Strait.

	r^2			RRMSE (%)		
Parameters	**Split half 1**	**Split half 2**	**LOO**	**Split half 1**	**Split half 2**	**LOO**
TDLP11	0.76	0.72	0.79	20.8	23.5	19.7
S/V	0.67	0.65	0.70	22.9	25.2	22.0
C/V	0.54	0.53	0.79	24.1	24.7	27.5
P/V	0.68	0.46	0.51	14.9	19.1	21.1
Ad/Al (V)	0.24	0.32	0.31	35.3	34.1	33.8
Af/Al (S)	0.24	0.26	0.69	29.8	28.6	20.6

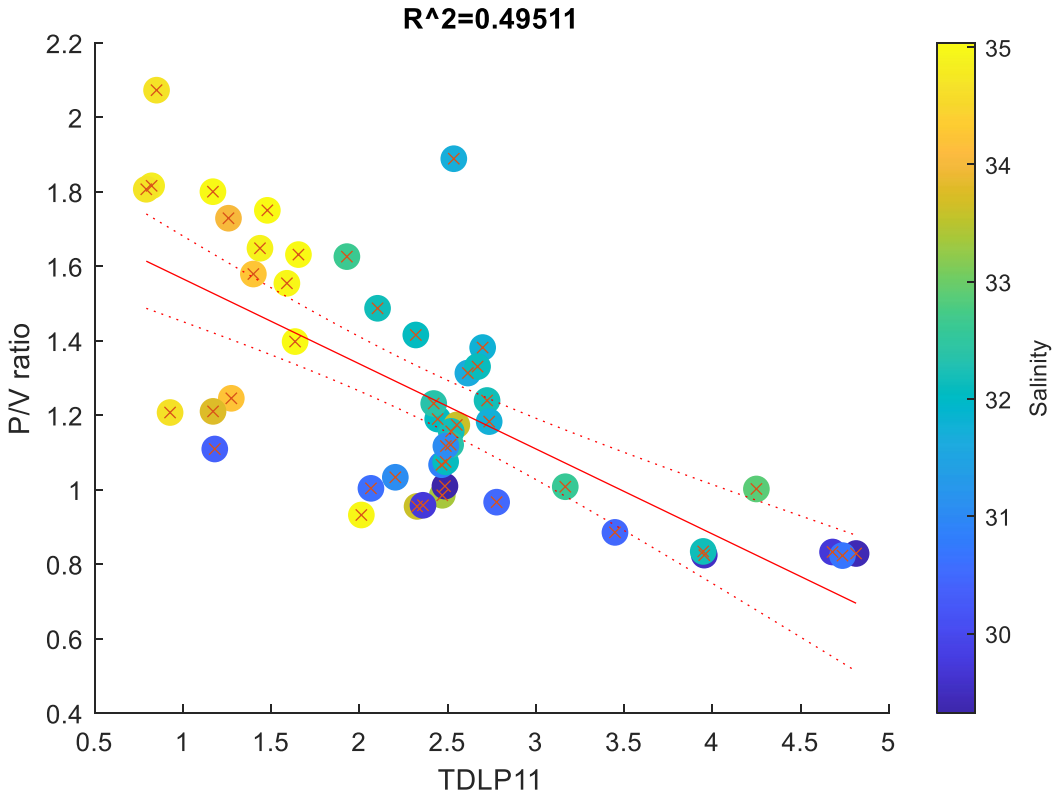


Figure S2: P/V versus TDLP11 coloured by salinity for the measured samples.

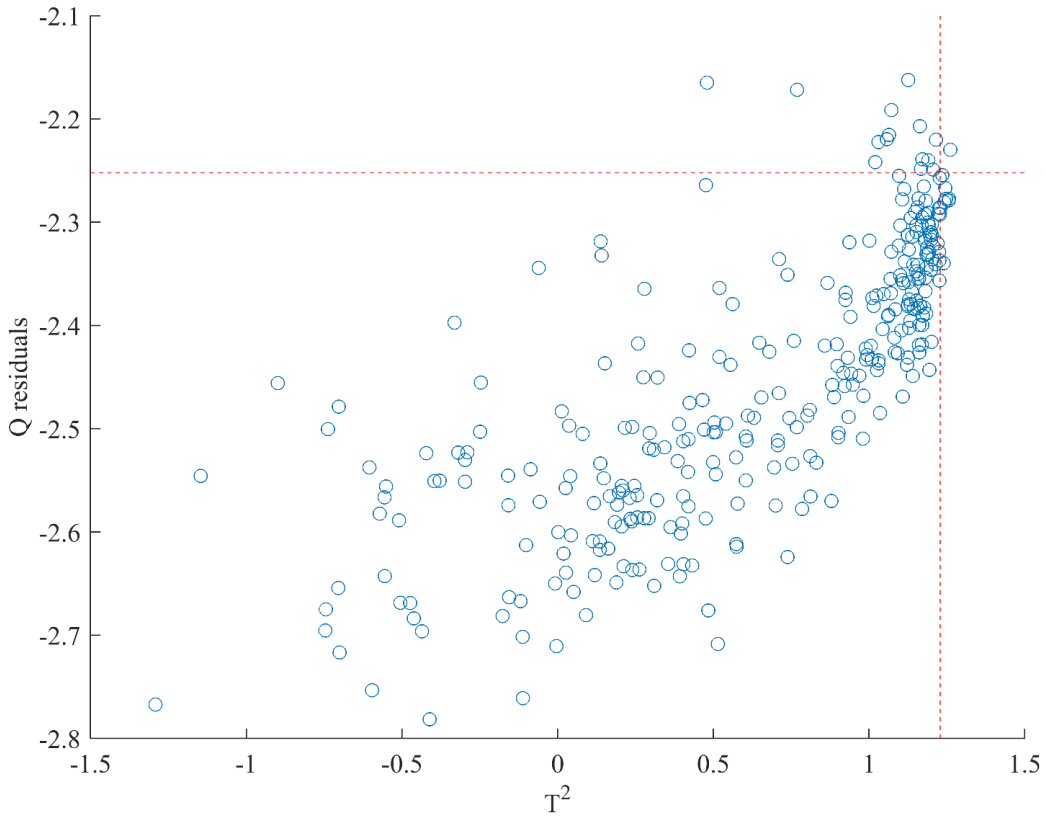


Figure S3: Hotelling plot for predictions of lignin phenol parameters using the EEMs measured across Fram Strait (n=313)

Table S2: Estimated p-values from ANOVA one-way test performed between the TDLP11 concentration and the diagenetic ratios of different sample groups, hence water masses, across the Fram Strait. A p-value < 0.05 means that the TDLP11 concentrations or the diagenetic ratios are significant different between the water masses compared. AW=Atlantic water, PSW=Polar surface water, MAW=Modified AW, NTSW=Norske Trough surface water.

Predicted								
Watermasses			**TDLP11**	**S/V**	**C/V**	**P/V**	**Ad/Al (V)**	**Ad/Al (S)**
AW	**vs**	**PSW**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
AW	**vs**	**NTSW**	<0.001	<0.001	<0.001	<0.001	1.00	<0.001
AW	**vs**	**MAW**	0.18	<0.001	<0.001	<0.001	0.11	0.11
PSW	**vs**	**NTSW**	<0.001	<0.001	<0.001	<0.001	0.29	0.98
PSW	**vs**	**MAW**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
NTSW	**vs**	**MAW**	<0.001	0.11	0.20	0.07	0.45	<0.001
Measured								
Watermasses			**TDLP11**	**S/V**	**C/V**	**P/V**	**Ad/Al (V)**	**Ad/Al (S)**
AW	**vs**	**PSW**	<0.001	<0.001	<0.001	<0.001	0.61	0.37
AW	**vs**	**MAW**	0.92	<0.001	<0.001	0.52	0.01	0.08
PSW	**vs**	**MAW**	0.86	0.16	0.78	0.31	<0.001	1.00

Table S3: Comparison of prediction power between N-PLS models built on full EEM and reduced EEM (four emission slices)

	r^2		RRMSE	
	Full EEM	**Reduced EEM**	**Full EEM**	**Reduced EEM**
TDLP11	0.76	0.75	20.8	21.7
S/V	0.67	0.67	22.9	22.8
C/V	0.54	0.56	24.1	23.5
P/V	0.68	0.63	14.9	15.8
Ac/Ad (V)	0.24	0.26	35.3	34.9
Ac/Ad (S)	0.24	0.28	29.8	28.9

Technical
University of
Denmark

DTU Aqua
Kemitorvet
DK-2800 Kgs. Lyngby

www.aqua.dtu.dk