Revisions

REPORT #1

My major comment was mentioned previously, but still only addressed on a qualitative, rather broad level:

Comment: Can the sharp decline in DOC and humic/aromatic CDOM markers also be explained by a skewed sampling of endmembers which have high concentrations but low volumetric input into the beach STE? This could be discussed (or at least mentioned) using rough estimates from other permafrost sites and/or Kipp on fluxes and residence times. In addition, how much DOC could have been removed by the reactive Fe on the sediments? Based on the available literature, Fe-DOM precipitation and microbial remineralization cannot in my view suffice to explain the sharp decline. Does DOC decline with DIC increase? How do oxygen, DIC and DOC behave in relation to each other? An attempt to try and close the carbon budget would be highly appreciated.

Response: We thank the reviewer for this thoughtful and constructive comment. We agree that constraining the relative importance of processes driving the sharp decline in DOC and CDOM, as well as attempting to close the carbon budget, would provide valuable insights. However, we respectfully emphasize that such quantification is not possible with the current dataset. Our study was specifically designed to resolve concentration gradients and compositional changes along the supra-permafrost flow path, rather than to quantify fluxes or establish a full carbon budget. In response to this comment, we have revised the title of the manuscript to better reflect the scope and limitations of our study. Specifically, we replaced the term “fate” with “qualitative transformations” to clarify that our work is based on concentration patterns and compositional changes, rather than on quantitative flux estimates or a closed carbon budget. This change is intended to improve the clarity of the manuscript and to better align reader expectations with the nature of the data and analyses presented.

Regarding the potential bias associated with endmember sampling, we acknowledge that meltwater and massive ice samples exhibit high DOC and CDOM concentrations but may represent limited volumetric contributions to the overall groundwater flux within the beach subterranean estuary (STE). However, our dataset also includes groundwater samples collected within the beach that, depending on tidal conditions, correspond to nearly 76% freshwater endmembers. These samples are characterized by near-zero salinity, strongly depleted isotopic signatures ($\delta^{18}O$ and $\delta^{2}H$), and elevated radon activities, indicating a dominant contribution from supra-permafrost groundwater with minimal to no mixing with seawater. These observations support the conceptual model presented in Fig. 2, where portions of the beach groundwater system can be dominated by direct inputs from thawing permafrost (including both active layer and potentially underlying massive ice), before significant mixing within the STE depending on the tide level. Therefore, we argue that the observed sharp DOC and CDOM gradients cannot solely be attributed to a sampling bias toward high-concentration but low-volume endmembers.

Importantly, in a complementary study (Kipp et al., 2025), we aimed to trace the contribution of massive ice meltwater using radioisotopic tracers to ultimately quantify groundwater-derived fluxes. However, our results indicate that beach groundwater does not simply reflect a

conservative mixing between upstream permafrost-derived waters (massive ice and active layer) and seawater. Instead, the radio isotopic signature of massive ice “modified” in beach groundwater, limiting its use to quantify submarine groundwater discharge. Although seawater recirculation occurs in the STE, it remains limited by the low tidal amplitude in the study area. Using stable isotope signatures, a first-order estimate of the relative contribution of fresh groundwater to beach groundwater samples has been derived using a simplified two-endmember mixing approach. The contribution of fresh supra-permafrost water ranges from approximately 20 to 76 %. Although these estimates do not constrain the total water flux through the beach system, they indicate that freshwater inputs can represent a substantial fraction of beach groundwater. As a result, beach groundwater constitutes a dynamically evolving endmember, rather than a simple mixture of predefined sources. This complexity, together with the rapid transformation and attenuation of initial source signatures, prevents a robust quantification of the relative contributions of each endmember and their associated fluxes.

We thank the reviewer for raising these important points regarding the potential role of Fe–DOM interactions, microbial remineralization (or reprocessing), and carbon transformations along the flow path. We agree that quantifying the amount of DOC removed through Fe–DOM precipitation and microbial processes would be highly valuable. However, such quantification is not possible with the current dataset, as it would require constrained estimates of water fluxes, reaction rates, and residence times, which are not available in this study. While DOC broadly follows the theoretical mixing line, a350 and HIX deviate below it in beach groundwater and seawater (Fig. 5B, 5C). The a350 values, as a proxy of the DOM concentrations, are also not conservative in the DOC pool (Fig. 6). This divergence is not consistent with simple dilution, which would be expected to affect all dissolved constituents proportionally. In addition, paired DIC, DOC, and O2 measurements show that beach groundwater exhibits median DIC concentrations exceeding 3,000 µmol L^{-1} and mean O2 saturation of ~34%, markedly different from seawater at comparable salinities. These patterns provide independent geochemical evidence consistent with microbial remineralization within the STE.

Regarding the role of reactive Fe, our measured Fe:DOC molar ratios suggest that Fe–DOM interactions likely operate primarily through molecular fractionation, selectively scavenging aromatic, high molecular weight humic compounds rather than inducing bulk DOC removal (Linkhorst et al., 2017; Riedel et al., 2013; Amoako et al., 2025). This interpretation is consistent with the deviation of a350 and HIX below the mixing line, while DOC exhibits a more conservative behaviour. Taken together, these observations support the view that the STE functions as a zone of DOM transformation, where microbial activity and Fe–mineral interactions contribute to compositional changes in the DOM pool, while mixing processes primarily govern bulk DOC behaviour.

We have clarified this positioning throughout the manuscript, and Sections 3.2, 3.3, and 3.4 have been revised accordingly to (i) evaluate the contribution of fresh supra-permaforst water contribution in STE based on stable isotope signatures of water, (ii) acknowledge the limitations related to endmember representativity and mixing processes, and (iii) present a more cautious interpretation of the mechanisms driving DOC and DOM changes. We also explicitly state that a formal carbon budget closure goes beyond the scope of this study, and we identify this as an important direction for future work.

Minor comments:

The citations are still wrong or in the wrong order (sometimes alphabetical, sometimes chronological) throughout parts of the manuscript. Correct for typos and/or order lines 89, 218, 222, 244, 353.

Response: We sincerely apologize for this oversight and for not having corrected all citation formatting issues in the previous revision. We have carefully reviewed the full manuscript again, including the lines identified by the reviewer (89, 218, 222, 244, and 353), and corrected the citation order and typographical inconsistencies throughout. We hope that all citation formatting issues have now been resolved in the revised version.

Some phrasing is unclear. E.g. line 105, what is a "lateral segregation of molecules"? Or lines 244-245 "oxygen input (…) is rapidly consumed". Or line 404 "protein-like compounds that are ultimately dilute…"

Response: line 105: We agree that the term "lateral segregation of molecules" was unclear. It has been replaced with "selective transport and fractionation of organic molecules along the flow path" to better reflect the processes described. Line 244-245: The sentence has been revised to clarify that oxygen supplied through tidal exchange is rapidly consumed by heterotrophic microbial processes. Line 404: The sentence has been revised to clarify that these compounds are diluted during mixing with adjacent coastal waters.

Line 114: FDOM abbreviation was not explained.

Response: We have now defined the abbreviation at its first occurrence in the manuscript. The sentence has been revised to read: "…including chromophoric dissolved organic matter (CDOM) and fluorescent dissolved organic matter (FDOM)…"

Lines 140-151: "similar geomorphological characteristics" to what?

Response: We have clarified that the geomorphological characteristics refer to those described for Peninsula Point. The sentence has been revised accordingly. "Crumbling Point (N = 5), located at the northwest edge of Kugmallit Bay, is another retrogressive thaw slump system with geomorphological characteristics similar to Peninsula Point."

Lines 152, 159, 160: How far away really is seawater sampling? 500-1000m offshore is not nearshore, but "in front of the slump system" seems to be a bit closer (how close?). Please explain.

Response: Seawater samples were collected approximately 0.5–1.0 km offshore from the coastline, along the seaward extension of the slump systems where terrestrial samples were collected. We have clarified this distance in the manuscript to avoid confusion.

Line 191: home-made standard or a certified reference material?

Response: The standards used were certified reference solutions: the EMSL secondary standard and the DSR Batch 21 Lot 04-21 from the University of Miami. This has been clarified in the manuscript.

Line 210: Explain the specifics of aCDOM350 and what it is a proxy for. Aromaticity? Terrestrial DOM? How is it different from SUVA254? The rationale for using it is not clear.

Response: We clarified in the manuscript that a350 is used as a proxy for CDOM concentration, while SUVA254 reflects DOM aromaticity normalized to DOC concentration. We revised the text to clarify the distinction between these two indices.

Lines 291-293 are not related much to the remainder of the paragraph/chapter.

Response: We agree and have moved this paragraph in Section 3.4.

Lines 344-345: FI and BIX are not reported but discussed? Why? In contrast to the statement earlier in the manuscript, they do seem to diverge between endmembers.

Response: FI and BIX showed limited variability and did not provide strong discrimination between sample types. We therefore decided to remove them from the manuscript and focused on the most informative indices (aCDOM350, SUVA254, and HIX).

Tables 1 and 2 have very different styles. Please unify.

Response: The style of the two tables has been unified

Figure 6: What is a global relationship?

Response: The term “global relationship” has been replaced with “relationship across all samples” to clarify that the relationship refers to all data points pooled across sites and sample types.

Figures 4 and 6 have different symbols for the endmembers meltwater and beach groundwater compared to Figures 5, 7 and 8. The inlets in Fig. 5 are again different (beach groundwater as blue squares opposed to yellow circles) and also the axes fonts need to be bigger.

Response : This was changed accordingly.

Report #2

Comments to Author and Editor

I've briefly checked all communications in prior rounds of review. I think the authors have made substantial efforts to moderate their claims, clarify terminology, and remove the most problematic elements (notably the Fe-spiking experiment). The revised manuscript is clearer and more cautious than the previous version.

However, I feel the central evidentiary limitation remains: the dataset still does not sufficiently and clearly discriminate between biogeochemical processing and conservative mixing as the dominant explanation for the observed DOM and DOC patterns in beach groundwater. The uncertainty around endmembers must also be acknowledged more explicitly. The revision improves interpretation, but it does not materially strengthen the causal inference. This distinction is important for the Editor's decision.

Below are my answers to the specific questions raised by the Editor.

*) Very generally speaking, reviewers felt that the dataset simply did not provide enough evidence for the fairly strong claims.

-> Claims in the Abstract and Conclusions were explicitly softened to frame beaches as potential rather than demonstrated reactors or sinks. The authors now explicitly state that whether Arctic beaches act as filters, sinks, or transient reactors remains an open question, which is an appropriate correction. Removal of the Fe-addition experiment eliminates an over-interpreted line of evidence that previously overstated mechanistic certainty.

Despite moderation of language, the interpretive structure of the paper still leans toward process attribution (microbial degradation + Fe–OM interactions) without explicit data or quantitative constraints on residence times or endmember variability. Key statements such as (Abstract) "microbial degradation and mineral–organic interactions jointly contribute to the degradation and retention of supra-permafrost groundwater DOM" and (Conclusions) "The findings highlight the importance of microbial degradation and mineral-organic interactions in transforming and retaining supra-permafrost groundwater DOM pool." are plausible, but not uniquely supported by the presented data. This is acceptable only if the Editor is satisfied with a hypothesis-generating contribution.

*) Was simple mixing (dilution) of multiple water sources considered strongly enough as causing the observed patterns? In particular when considering sampling bias affecting endmembers?

-> No, I don't think mixing alone could explain all observed patterns in the data. For instance, the high variability in DOC concentrations and CDOM values at low salinity ranges indicate the presence of other input/loss processes. This interpretation, however, implicitly assumes that the four sampled water types (massive ice, meltwater, beach groundwater, and seawater) are hydrologically connected, which I only assumed so. This is I think reasonable given the isotopic and salinity data.

The authors explicitly show that DOC follows a conservative mixing line between freshwater and saline endmembers once groundwater enters the STE. Stable water isotopes ($\delta^{18}O$, $\delta^{2}H$) and

salinity are correctly used to demonstrate binary mixing between supra-permafrost water and recirculated seawater. The manuscript now clearly states that dilution dominates DOC behavior in seawater, which is an improvement.

However, beach groundwater is still treated implicitly as a transformed pool, rather than first being treated as a mixture whose composition must be quantified. There are also no volumetric constraints on meltwater production, groundwater discharge, seawater recirculation rates, or residence times (despite the availability of radon data in Kipp et al.). In addition, and more importantly, endmember uncertainty is not propagated: for instance, meltwater samples are few and variable, and seawater is strongly influenced by the Mackenzie plume.

Having said that, it is important to clarify that volumetric contributions of endmembers are not required. For qualitative indices (e.g., SUVA and HIX), DOC concentrations alone are sufficient to evaluate mixing behavior. For quantitative indices such as DOC concentrations, CDOM absorption, and fluorescence intensities, only the absolute values of the endmembers and samples are required. Under the assumption of conservative behavior and well-defined endmembers, relative contributions can, in principle, be estimated using a simple two-endmember mixing model based on these quantitative indices.

So, for instance regarding the comment from another Reviewer "There are no volumetric estimates for meltwater production to estimate if the DOM-rich meltwater is able to leave a traceable signal in the beach groundwater inventory", if groundwater samples had significantly (which needs statistical testing) higher values than the low-value endmember (seawater in most cases), then this alone indicates a detectable influence from meltwater. The main limitation here is not the absence of volumetric data per se, but the insufficient characterization and uncertainty of the endmembers themselves.

Response:

We would like to thank the editor for sharing these additional comments, which were helpful in better understanding the concerns raised during the review process. At the same time, we note that some of these comments appear to originate from internal or cross-reviewer discussions. While we appreciate the transparency, we understand that such exchanges are not always intended to be shared with authors, and we leave it to the editor's discretion in this regard.

We recognize that our study addresses processes that remain poorly documented in Arctic coastal environments, and that this may raise important questions and uncertainties. We believe that this reflects the novelty of the work, which aims to provide new insights into supra-permafrost groundwater flow and DOM transformations in Arctic beach systems. We acknowledge that the dataset has inherent limitations, as discussed throughout the revised manuscript. However, we are confident that the study offers a valuable and original contribution to the field, particularly given the scarcity of observations in these environments.

Major comments
Aside from the points above, I personally have serious concerns on the quality of PARAFAC. First, it's obvious that C3 has two emission peaks, which is not theoretically possible. Second, three components are too few to for a valid PARAFAC model of DOM dataset. As hinted in Murphy et al. (2018) "Photochemistry illuminates ubiquitous organic matter fluorescence spectra" and mentioned in Wünsch et al. (2019) "Emerging patterns in the global distribution of dissolved organic matter fluorescence", the adequate description of DOM fluorescence requires at least three components emitting in the visible range. With more than 250 samples, you should be able to obtain more components. In more careful modeling, I expect C3 will be separated into one protein (tryptophan)-like component and one humic-like component, and additional humic-like component should also emerge to make the total humic-like components to at least three.

Response: We thank the reviewer for this insightful and constructive comment regarding the PARAFAC model. In the previous version of the manuscript, we chose to present a 3-components model as these components explained the vast majority of the variance and represented the dominant fluorescent signals in the dataset. Additional components provided limited added value and did not significantly improve the overall interpretation. However, we do agree that having two emission peaks is not theoretically expected. In response, we revisited the PARAFAC analysis and updated our model with a 5-component one, which was validated using standard procedures, including split-half analysis. As suggested by the reviewer, the previously identified C3 component was resolved into one protein-like (tryptophan-like) component and additional humic-like components. This resulted in an improved separation of fluorescence signals, providing a more detailed characterization of DOM composition. The revised model now includes two terrestrial humic-like components (C1 and C4), one microbial humic-like component (C5), and two protein-like components (C2: tyrosine-like; C3: tryptophan-like; Table 1). We note, however, that the additional components (C3-C5) contribute relatively little to the total fluorescence signal and do not exhibit patterns that differ substantially from those already captured by the original (previous) model.

Therefore, while the 5-component model refines the description of DOM composition and is more consistent with the theoretical FDOM framework, it does not alter the study's main interpretations. has been revised according to incorporate this updated model and to clarify the methodological approach and its limitations.

Comment: Also please indicate if an instrument-specific bias correction has been made. If not, your fluorescence results are not comparable to others with appropriate corrections.

Response: Thank you for raising this point. Fluorescence spectra were corrected for instrument-specific excitation and emission bias using manufacturer-provided correction files implemented in the staRdom workflow (eem_spectral_cor() function; Pucher et al., 2019). This clarification has been added to the Methods section.

The text can now read (line 202): "Fresh deionized water was used as a blank and absorbance measurements were used to correct the inner-filter effect. Fluorescence spectra were further corrected for instrument-specific excitation and emission bias using manufacturer-provided

correction files, and the dataset was corrected for Rayleigh and Raman scattering, according to the method described by Pucher et al. (2019)."

Figure caption of Fig. 5 "stable isotope signatures" Please indicate which stable isotopes you mean here and which sample types were chosen as endmembers (meltwater and seawater? Or groundwater and seawater?). You should also clearly state this in M&M regarding how you defined the endmembers and drew a conservative mixing line (maybe I have missed it).

Response: The conservative mixing lines in Fig. 5 were constructed using the meltwater samples as the freshwater endmember and the most saline seawater sample as the marine endmember, with salinity as the conservative tracer. This has now been explicitly stated in Section 2.3, and the text now reads: "To evaluate conservative mixing behaviour, a two-endmember mixing line was constructed using the most saline seawater sample as the marine endmember S = 20.08 (N = 1) and the meltwater samples as the freshwater endmember S = 0.18-0.30 (N = 8), with salinity as the conservative tracer."

Minor comments:

Comment: Since Fe-spiking experiments have been removed, its description should also be removed from PANGAEA summary of the dataset.

Response: This has been done and the data has been resubmitted.

Comment: aCDOM350 should be a350, which is the common terminology in the filed.

Response: This was corrected

L21 in DOC and chromophoric DOM (CDOM) -> in DOC and chromophoric DOM (CDOM) with respect to/than expected from conservative mixing?

Response: We have clarified the text to specify that the observed decrease is relative to values expected from conservative mixing. The revised sentence now reads as: "…we observed sharp declines in DOC and chromophoric DOM (CDOM) relative to conservative mixing, indicating substantial removal along the short-distance flow path."

L24/526 Was this correlation between PARAFAC components calculated with absolute intensities (RU) or relative intensities (%)? If the latter, this has no meaning because they sum up to 100% and there is a strong negative correlation bias. Now I've noticed that there is no stats section in M&M. Please make it and explain all statistical analyses to clarify what you actually did.

Response: Correlations between PARAFAC components and optical indices were calculated using absolute fluorescence intensities (RU), not relative contributions. This has been clarified in Section 2.4 of the revised manuscript, which now includes statistical clarifications. The text can now read: "Pearson correlation coefficients were calculated between PARAFAC components, expressed as absolute fluorescence intensities (RU), and between components and optical indices. Correlations were considered significant at $p < 0.001$. Differences in PARAFAC component intensities across

sample types were assessed using Kruskal-Wallis tests followed by pairwise Wilcoxon tests with Bonferroni correction."

Statistical results have been reported where significant: DOC concentrations differed significantly across sample types (Kruskal-Wallis, $p < 0.001$; Section 3.3), and C2 absolute intensities differed significantly across sample types ($H = 10.21$, $df = 3$, $p = 0.017$), with pairwise comparisons revealing significantly higher values in seawater relative to meltwater (Wilcoxon test, $p = 0.013$, Bonferroni-corrected; Section 3.5).

Comment: Indicate how you confirmed accuracy in TOC analysis by the CHNSO analyzer (isn't it just precision?). Also, what is "analytical uncertainty" in AES and Shimadzu TOC analysis, respectively? Is it precision or trueness, and how you checked it? Is it different from "reproducibility" for the Fetot analysis? Please use correct and consistent terminology. Note that accuracy is an integrated concept of precision and trueness.

Response: Analytical terminology has been revised and corrected throughout Section 2.3: the values reported for the CHNSO analyzer and AES now refer explicitly to analytical precision; trueness for the Shimadzu TOC analyzer was verified against certified reference solutions and has been clarified accordingly; and the term "accuracy" has been replaced throughout to reflect the distinction between precision and trueness. The term "reproducibility" used for the Fetot analysis has been retained, as it correctly describes the consistency of repeated measurements.

L298 The unit of SUVA254 is not mgC L-1. It's L mgC-1 m-1, since it's absorbance (1-m path length) divided by DOC concentration. Also use subscript SUVA254 for all. Correct these in ms and Figs/Tables.5

Response: This has been corrected

L375 Fig. 2 -> Fig. 4?

Response: The reference to Fig. 2 was intended to direct the reader to the conceptual schematic illustrating the theoretical mixing model. To avoid confusion, we have replaced the reference with Fig. 4, which shows the stable isotope signatures and more directly supports the statement.

L423 Because SUVA (and HIX) is a qualitative index rather than quantitative (concentration-based), the mixing line should appear as a curved line rather than a straight line connecting the end-members, reflecting differences in their relative concentrations. Please recheck this.

Response: We agree that SUVA and HIX are ratio-based indices for which a conservative mixing line is not appropriate. We have decided to remove the mixing lines from these panels in Figure 5 since they do not affect our results.

L438 retained in where/what?

Response: The sentence has been revised to clarify that HMW CDOM compounds are retained within the STE sediments.

L445 I couldn't understand how you can say this ("particularly meltwater, massive ice, and beach groundwater samples") from Fig. 6. Perhaps adding a mixing line would help?

Response: Upon reflection, we agree that the figure does not clearly support the distinction between sample types in this context. The text has been revised accordingly and can now read as: "The strong decoupling between DOC and a350 (Fig. 6), indicates that these variables follow different trajectories along the supra-permafrost flow path (Fig. 5A and 5B), leading to a decrease in humification and a shift toward lower molecular weight material as revealed by HIX indices (Fig. 5A to 5C)."

L478 together with elevated SUVA254 values of seawater samples; compared to what? They were not higher than other samples.

Response: The text has been corrected; this was already flagged by Reviewer #1.