

Overall, the manuscript presents an impressive and very detailed dataset of transects in various fjord branches with very different biological and biogeochemical characteristics. The focus on different glacial bays and the large variety of data provided are very valuable.

Thank you for your appreciation and great expertise with numerous helpful suggestions for the substantial improvement of the manuscript.

However, the manuscript appears somewhat unfocused, overly lengthy, and too descriptive. I strongly suggest a more focused, hypothesis-driven approach, utilizing only the data and analyses necessary to test the hypotheses, and putting some of the plots in the supplement. The results and discussion sections are currently very long and dense, making the narrative difficult to follow, and some plots are challenging to interpret. The station examples appear very descriptive.

We agree with the reviewer that the manuscript may seem too long and too descriptive. It is always a challenge to make the simple conclusions out of the multidisciplinary study with plenty of parameters measured and to summarise it in a condense form without losing the huge complexity.

And indeed, in our case the complexity is the main character, because what we want to show is that none hypothesis works at all places, each system is unique. Therefore, we presented the station-based examples of small-scale gradients, to exemplify how variable individual system are locally.

What we can do for improvement is to focus on the most interesting observations and to arrange it across the studied spatio-temporal factors and how important they are for which parameter and why. This will be arranged around the questions: Which environmental parameters are the most heterogeneous, where is that heterogeneity greatest, and which factors drive it.

To do that, first, the heterogeneity within each parameter and studied factor can be analysed by the analysis of the coefficient of variation (CV) as a better measure than SD because it normalizes by the mean.

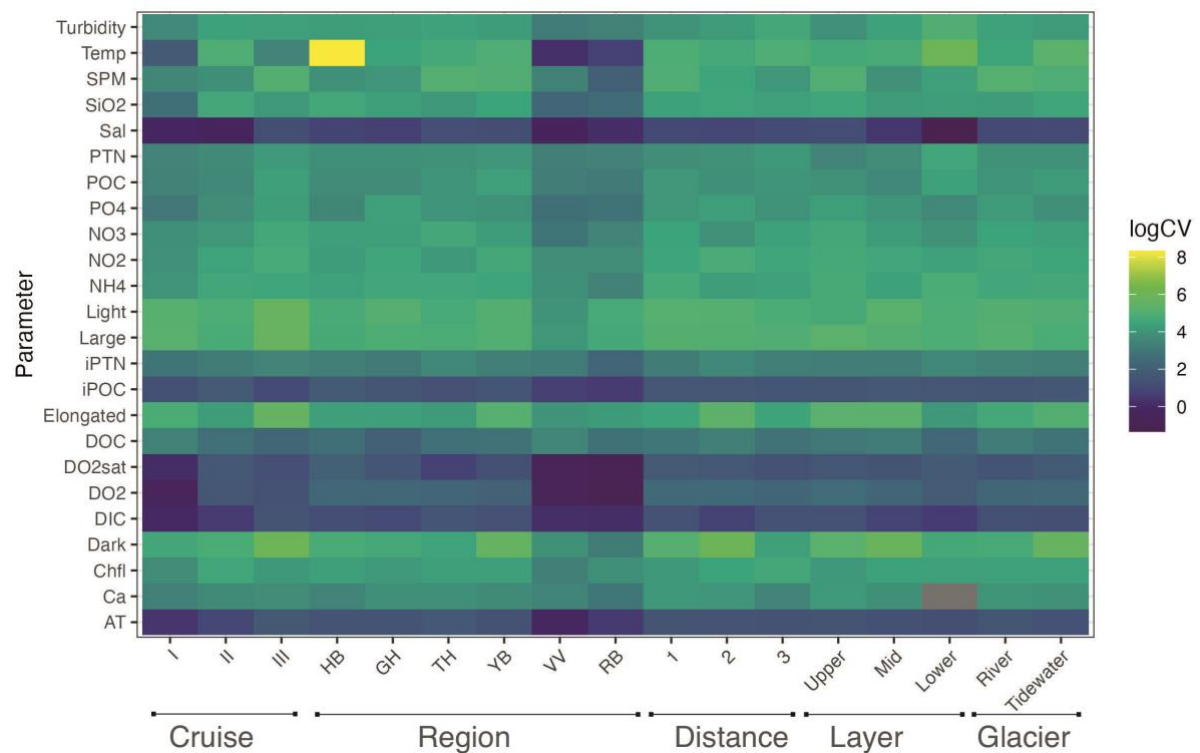
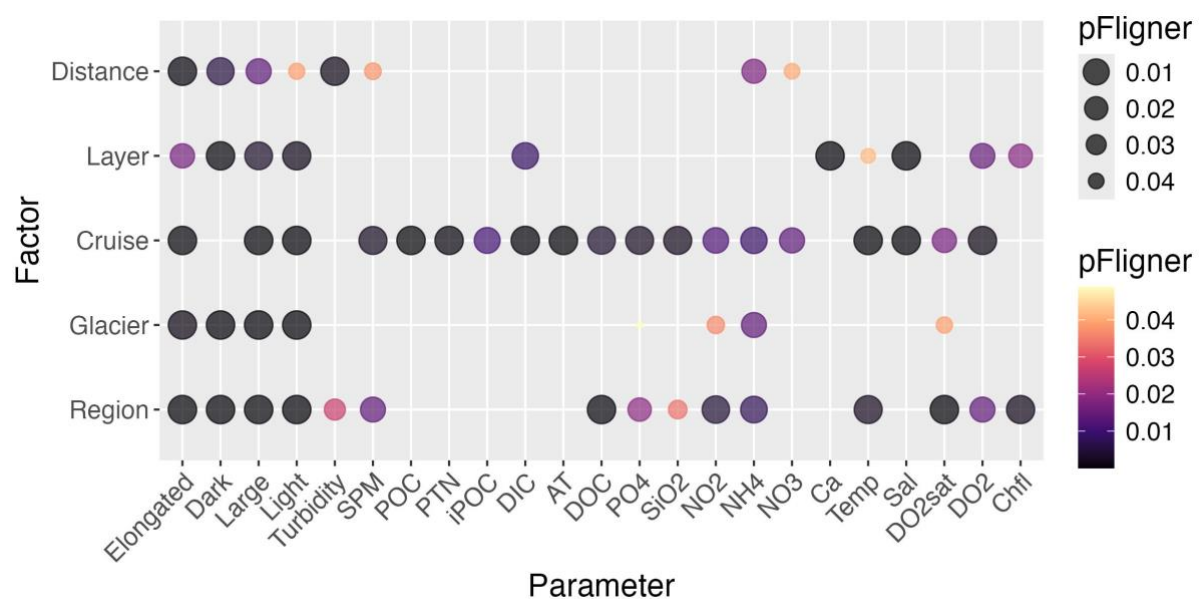
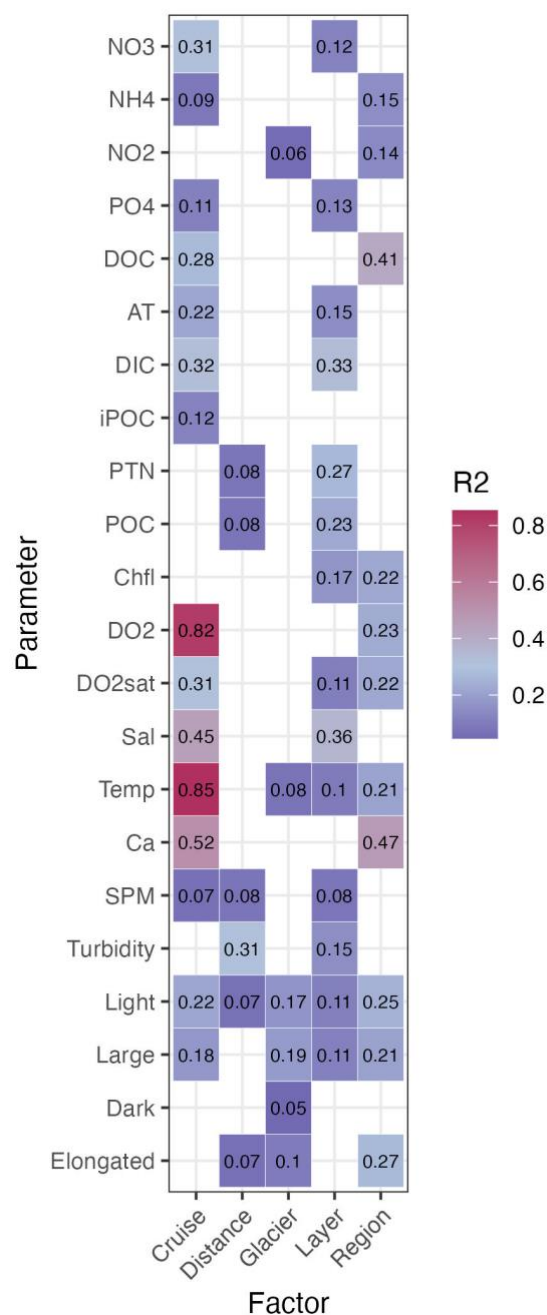


Fig. S11. Heatmap of local heterogeneity: the coefficient of variation (CV) across the studied environmental parameters and factors.

Then, to test among which factors variability differs significantly, we performed the Flinger tests, which showed that for most of the studied parameters temporal effect was significant (Cruise), and that all of the studied factors were significant for marine snow morphotypes. Distance from the glacier was additionally significant for Turbidity, SPM, NH₄, and NO₃. The water column stratification was also significant for temperature, salinity, oxygen concentration, chlorophyll a, and DIC. Glacier type was significant also for NH₄, NO₂, and oxygen saturation.



And finally, we performed ANOVA and linear model as the statistical approach to estimate the Effect Sizes to answer which factor explains most of the variability and how much, which can be visualized by the R^2 :



The hypothesis of extreme spatiotemporal variability on very small scales is compelling and could be crystallized further. Right now, the authors seem to attempt to include everything, including small-scale examples, which makes it difficult to maintain an overview. The Redundancy Analysis (RDA) appears to be the core of the paper and its most useful component; it might be worth centering the manuscript around these findings.

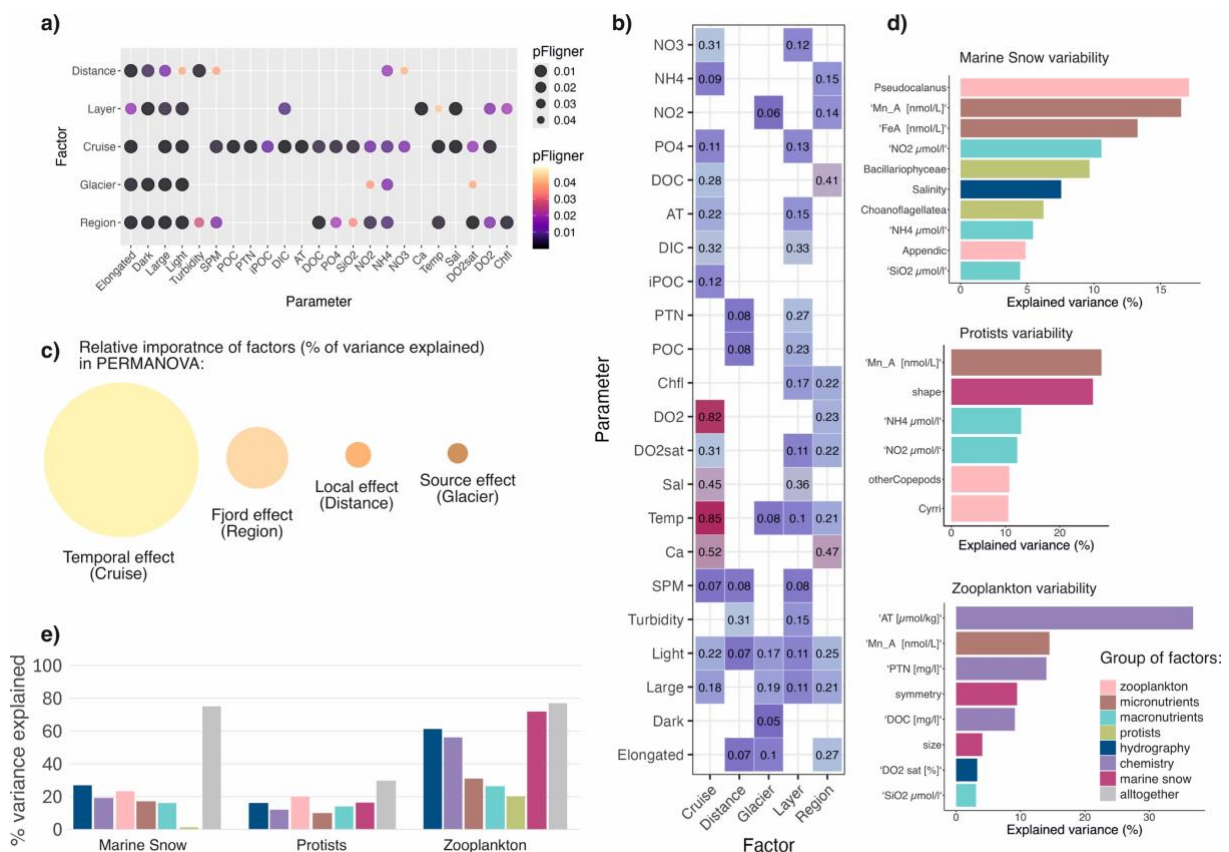
Thank you for this suggestion. We think we can replace the lengthy description of each case by focusing only on most significant relations and observations. As showed above, we can also add more statistical tests to strengthen also the parameter-level interpretations which should facilitate more robust and general than descriptive approach to results.

However, the presentation of the figures and methods for the RDA is not entirely clear. The statistical methods could be more detailed, specifically by clarifying which are the response and which are the explanatory variables. It also seems that the RDA is overparameterized, contradicting the methods section which mentions that only a subset of drivers was selected. Many arrows lack labels, and it is unclear how the most influential variables were selected or defined.

We agree that the RDA method needs more explanation and simplification. It will be greatly improved. The holistic database consisted of 56 parameters, including hydrography (temperature, salinity, oxygen saturation, SPM, chlorophyll a), macronutrients (NH₄, NO₃, NO₂, PO₄, SiO₂), sediment-bound micronutrients (FeD, FeA, MnA, MnD), biochemistry (AT, POC, DOC, DIC, PTN, δ¹³C and δ¹⁵N), marine snow (concentrations of 4 morphotypes, total abundance of particles, their morphological traits: lightness, size, shape, symmetry, heterogeneity, structure), protists (11 taxonomical categories), and zooplankton (10 taxonomical categories) at 40 stations that had a complete set of measurements.

Additionally, because too many predictors relative to samples and response variables were available, the number of environmental variables was reduced by removing the redundant factors that were highly co-correlated with others. After the forward selection, the step model was run on only significant predictors for which bar plots were made to present the percent of variance explained by factors selected as the most significant.

This Figure can also be modified to include also other analyses than RDA:



Furthermore, I am not convinced that it is possible to discuss "seasonality" based on only two time points in spring (I am not convinced it is a pre-bloom state) and summer. It might be more accurate to describe this as "temporal variability."

Fully agree. This will be changed accordingly throughout the manuscript.

The English language would benefit from careful proofreading, as there are several awkward formulations and typos. I provide some examples below.

The manuscript will be sent to the professional language editing service before the re-submission.

Additionally, the references in the introduction and discussion are often long, exhaustive lists. I suggest focusing on key, selected references rather than citing every study that has performed similar work.

Our intension was to show that despite the long list of good studies, our knowledge is still limited and we think in those few places it is important to show that the review of the topic has been done and still the knowledge gaps can be identified.

Finally, there are several formatting issues: taxa should be italicized, completed, and corrected (e.g., Cirripedia was spelled incorrectly). Subscripts and superscripts must be used consistently, and figure text needs to be large enough to be legible.

Thank you for pointing out those editorial shortcomings. We will carefully address and correct them.

Specific Comments:

L32: The statement that no two fjords have the same drivers seems oversimplified; some drivers are likely universal (e.g., temperature).

Agree, it may be modified by e.g. "No two fjords are the same, nor are the roles of the driving forces and main mechanisms that shape them."

L60: Do some of these autonomous vehicles take physical samples as well?

Some yes, but in a very limited amount and capacity and without the vertical resolution – mostly only surface is sampled.

L65: There are numerous seasonal studies, especially on hydrography and phytoplankton. I disagree with the assertion that this is a gap; perhaps specify if you are referring specifically to seasonal marine snow studies?

We agree and thus we can change the argumentation to "Resolving also the seasonal dynamics in fjords is important because different seasons present sharp changes in biological productivity"

L77: Typo: "faecel" should be "fecal."

OK.

L100f: This is a very long and awkward sentence with grammatical issues, an example of where professional proofreading would improve the manuscript.

Agree and indeed we intend to send the text to professional proofreading after the preparation of the modified, final version of the manuscript and before the re-submission.

Fig 1: Great figure. However, why were no reference stations sampled outside the fjord bays? This might be worth discussing or justifying.

The study focuses on the local impact of glaciers on the fjords' biogeochemistry, whereas the situation outside the fjords is very different and shaped by very different oceanographic conditions.

L117: More details about the glass bottles would be helpful. Were they acid-washed? Did you use a crimper to ensure they were gas-free?

The samples were collected to the precleaned, acid-washed (3M HNO₃) and 3 times MQ water-rinsed 250 ml glass bottles. To prevent air bubbles and contact with air, samples were poured into the bottles from the bottom (with the sampling tube inserted near the bottom) such that at least 100% of the bottle's volume overflowed. After filling, the samples were preserved with 100 µL of saturated HgCl₂ and tightly sealed using grease, a ground-glass stopper, and a retaining clip. They were then stored at +4°C until analysis.

L119: Should "AT" be "TA" (Total Alkalinity)?

Yes, it will be corrected accordingly.

L188: Did you use a binocular or a stereo microscope?

Thank you for pointing this out, it was stereo microscope.

L236ff: As noted above, I am not sure it is possible to discuss seasonality based on two time points in late winter (post-spring bloom) and summer. "Temporal variability" would be more accurate.

Agree. It will be changed accordingly.

L245: Typo: "Cyrripedia" should Cirripedia. I suggest double-checking all taxonomic names in the manuscript for spelling and ensuring they are in italics.

Yes, we definitely will correct all the spelling mistakes.

L257: Awkward phrasing. It is unclear which year this "low O₂ high Chl" observation refers to.

Agree, it can be modified as e.g., "Differences between two August cruises were observed in terms of the very low oxygen saturation, the highest chlorophyll concentrations, the lowest δ¹³C, high DOC, and much elevated levels of FeA and MnD in 2022 compared to 2023."

Fig 3C: Please use full, italicized taxonomic names and verify their accuracy (e.g., *Cirripedia*).

OK.

Fig 3: Some labels are too small to read. Units need to be in superscript (e.g., ind m⁻²)

OK.

Fig 4: Typically, units should use superscripts.

We are afraid they would not be well visible as superscripts, so we may keep them in [] brackets.

Fig 5: See comments for Fig 3.

OK.

L345: The "2" in O₂ needs to be in subscript.

OK.

Ch 3.3 and Fig 8: This seems to be the core of the paper. However, it remains unclear. Please detail the statistics in the methods (response vs. explanatory variables). The RDA appears overparameterized. Many arrows lack labels, and the criteria for selecting the most influential ones are unclear. Fig 8c is missing an x-axis label, making it difficult to interpret.

OK, we can better explain what was used in which statistics.

"To ask the question which parameter and where is most heterogeneous, the coefficient of variation (CV) was calculated for each parameter within each level of the studied factors: time (defined by cruise), region (fjords and bays), distance from the glacier/river source (stations coded as 1-2-3 along the studied gradients), depth layer (Upper (10-0m), Mid (50-10), Lower (>50m), and type of the glacier impact (Tidewater vs River). Then, to verify if the observed variances are significantly different across the studied factors, Flinger tests were performed. Further, to identify the main drivers of variability in each environmental parameter, separate ANOVAs and linear models were fitted for each predictor (Region, Glacier, Layer, Distance, and Cruise). For each environmental parameter, the coefficient of determination (R^2), adjusted R^2 , F-statistic, p-value, and effect size (η^2) were calculated.

Permanova was used on the multivariate holistic database of all the parameters scaled together, consisting of 56 parameters, including hydrography (temperature, salinity, oxygen saturation, SPM, chlorophyll a), macronutrients (NH₄, NO₃, NO₂, PO₄, SiO₂), sediment-bound micronutrients (FeD, FeA, MnA, MnD), biochemistry (AT, POC, DOC, DIC, PTN, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$), marine snow (concentrations of 4 morphotypes, total abundance of particles, their morphological traits: lighthness, size, shape, symmetry, heterogeneity, structure), protists (11 taxonomical categories), and zooplankton (10 taxonomical categories) at 40 stations that had a complete set of measurements.

Furthermore, partial Redundancy Analysis (RDA) was calculated on the same holistic database to test the effects of groups of individual on compositions of marine snow, zooplankton, and protists, and to test effects of each group of factors separately. In these models, we removed the temporal variation, by controlling the effect of cruise. Additionally, because too many predictors relative to samples and response variables were available, the number of environmental variables was reduced by removing the redundant factors that were highly co-correlated with others. After the forward selection, the step model was run on only significant predictors for which bar plots were made to present the percent of variance explained by factors selected as the most significant. First type of models was run separately on the compositions of marine snow, zooplankton, and protists as response variables, and the list of all remaining parameters in holistic database as explanatory variables. Second type of models was run separately on the compositions of marine snow, zooplankton, and protists as response variables and separately for individual groups of parameters (hydrography, chemistry, micronutrients, macronutrients, zooplankton, protists, marine snow) as explanatory variables, to verify how much of variation is explained when only specific subpart of the parameters is used."

L435: Is "paradigm" the right term here?

We agree the word paradigm was not appropriate. We would rather say "This study addresses an important knowledge gap by zooming in these complex systems and by investigating them through the lens of several studied factors."

L445: Please be careful discussing seasonality with only two time points; consider using "temporal variability" instead.

Agree.

L454: Would the decaying phytoplankton help explain the seasonal timing? It appears you may have sampled a post-bloom scenario.

Yes, we think so too. This part will be heavily modified but we will try to incorporate this argument in the new version of the temporal effects observed.

The data are consistent with ongoing spring succession and possibly with conditions following the diatom peak, but they do not demonstrate a uniform post-bloom state. Low silicate concentrations and low phytoplankton abundance may indicate preceding production, whereas subsurface chlorophyll peaks in several station profiles and oxygen supersaturation at depth suggest that primary production may still have been ongoing at the time of sampling. The abundant *Light* forms do not help distinguish between these possibilities, as they were identified solely as a morphotype based on the morphological features of UVP images, without direct analysis of their composition. As discussed in the manuscript, they may have represented freshly formed aggregates of decaying phytoplankton, but they could also have resulted from the flocculation of glacial flour during the mixing of meltwater with saline waters.

L465f: It is interesting to mention TEP, but since you did not measure it, perhaps frame this as an outlook for future research?

Agree, this part will be deleted.

L490: I am not convinced by your interpretation. Low silicate and high Cirripedia abundances point to a post-bloom scenario. In a pre-bloom stage, nutrients should not be depleted and zooplankton should not be abundant. Perhaps, justify with a key reference if you disagree.

We agree that a kind of inconsistency was introduced. We can clarify that as follows:

"The May observations, with cold, saline and highly oxygenated conditions (Fig. 4) most probably experienced the early phase of the spring plankton succession, when phytoplankton concentrations were still low (Fig. 3D) and the availability of nitrate and phosphate were still relatively high (Fig. 4). But the community composition and low silicate concentrations may reflect preceding diatom activity. Zooplankton was overwhelmingly dominated by the meroplankton (Fig. 3C), namely the Cirripedia - larvae of barnacles, which is typical for spring (Stübner et al., 2016; Weydmann-Zwolicka et al., 2021), while the high abundance of nauplii of copepods indicated that seasonal recruitment pulse already occurred, possibly supported by preceding or ongoing phytoplankton production (Gluchowska et al., 2025). Together, these observations suggest that spring production had already begun, although phytoplankton standing stocks were low at the time of sampling. The low protist concentrations may additionally reflect strong grazing pressure, one of the main controls on Arctic primary production alongside light and nutrient availability (Arrigo and van Dijken, 2015; Rysgaard et al., 1999)."

L569ff: This section is largely repetitive of the previous discussion.

Discussion will be fundamentally re-written.