Reviewer 3 (Sofia Ribeiro):

Harning and co-authors present a study of biomarker distribution in surface sediment samples from Baffin Bay, with focus on identifying biomarker signatures for the North Water Polynya that may inform paleo-environmental reconstructions in the region. This is both a well-justified and significant effort. On the one hand, the North Water is a globally important ecosystem that is particularly vulnerable to climate change. On the other hand, there is considerable uncertainty in the behavior, source producers (HBIs and sterols), and fidelity of biomarker proxies and thus local and regional data are essential to refine their usability.

The study includes a broad range of biomarkers and this is, in my opinion, the main strength of the work. It is also very well-written, and the figures are generally well prepared and adequate (see below for some suggestions for improvement).

My comments will focus on the HBIs and sterol work, as GDGTs are outside my realm of expertise.

While the study provides some useful insights (adding, for example, to the discussion on whether HBI III and HBI IV might in fact be produced by sympagic taxa), I found some of the data analyses and conclusions problematic and cannot recommend publication of the paper in its present form. I encourage the authors to consider a few points and revise the manuscript accordingly.

We greatly appreciate Sofia Ribeiro's time and consideration of our manuscript and thank the reviewer for a constructive critique that will lead to a stronger paper. Below we address each comment individually.

General comments:

- Limited number of samples and novelty

The study includes a surprisingly low number of samples (n=13, n=9 analyzed for bulk geochemistry) to represent such a large area. I found there are some overstatements throughout the manuscript that should be corrected having in mind the small sample size, and what indeed it adds in terms of novelty (or not) to the large Kolling et al. 2020 study. For example, in the Introduction it is mentioned that samples were collected in 2008 and 2017 but the total number is not given. From the table I assume 2008 (n=3), 2017 (n=10). It is also written in the introduction that biomarker proxies were assessed against modern instrumental data but this was only presented for GDGTs. Another overstatement, although minor, in the discussion (line 299) is saying that "several additional sterols" were added compared to the work of Kolling et al. 2020 when in fact only two more were analysed.

We fully acknowledge the limitation of our dataset and have tried to highlight this where relevant in the manuscript. Part of this study's novelty is that we include a broader assessment of biomarkers than traditional studies, as well as take advantage of large, previously published datasets (e.g., Kolling et al., 2020). For the latter, the authors presented HBI and sterol datasets to explore the utility of sea ice proxies in Baffin Bay, but not regarding the NOW specifically. We will edit the main text to remove and/or reduce any overstatements as pointed out here and highlight the novelty where needed.

- Lack of information on surface sediment samples

In order to be able to assess the new findings, it is important that the authors provide additional information on the samples. Table 1 can be expanded to include at least “year of collection” (might not be obvious to all from the cruise code), “coring device”, and data on bulk geochemistry. Figure 3 shows 13C and C/N data for the 9 samples, but it is not clear which ones these are. Also, TOC values should be given in the table, as these are important when considering biomarker concentrations (see comment below also). Is there any age control on the surface samples? Were any of these cores analysed for 137Cs/210Pb? This would give some confidence at least that they might indeed represent recent deposition.

We apologize for any information that we did not include that may be important for the reader. In the revised manuscript, we will add the year of collection, coring device, and indicate which samples were analyzed for bulk geochemistry to Table 1. Data for bulk geochemistry will be submitted along with the biomarker data to the PANGAEA online repository. Unfortunately, we do not have any quantitative age control on the surface samples, so we restrict ourselves to providing the thickness of the sample interval in centimeters. However, we do note that these samples all contain modern sediment as stained living foraminifera were present.

- Comparison with Kolling et al 2020

I found some of the comparisons with the Kolling et al. dataset quite confusing. In the materials and methods, it is written that the new dataset (n=13) will be compared with a subset of samples (n=70) from Kolling et al and that samples collected within fjords and bays will be excluded. However, later in the discussion, it is argued that the difference in brassicasterol and dinosterol trends between the two are likely due to the fact that some of the sites are in the vicinity of large fjords. If there is uncertainty whether some samples might be skewing the response, the authors could run the comparison with a different subset and evaluate if this is the case. Given that the Kolling et al dataset includes many more samples, the ranges of water depths, sedimentation rates, and likely sediment composition are likely larger than for the n=13 dataset. Simply comparing biomarker concentrations per volume of sediment across the two datasets and the now vs. non-now sites is not adequate, in my opinion.

Apologies for any confusion. We removed any sites from Kolling et al. (2020) from immediately within the fjords and bays. The sample sites that we mention regarding brassicasterol and dinosterol trends are not within the fjords/bays but located proximally within the main region of Baffin Bay. As suggested by reviewer 2, we did test whether removing these sites brings the two study’s datasets into closer agreement, however, that was not the case. We will investigate whether these differences may be the result of varying TOC content or not.

- Influence of TOC contents on the biomarker signals

Given that the NOW is a highly productive area, one can expect that TOC values for the NOW sites will be generally higher than for the non-NOW sites. It has been recommended, and is common practice in paleo sea ice reconstructions using HBIs, to normalize the data by TOC. This way, we account for down-core changes in sediment composition. The same would be important for a dataset like this one, where large changes in sediment composition and organic matter content can be expected across the region. I strongly recommend that the authors plot all biomarker data in ng.gTOC-1 and revise the discussion and conclusions accordingly. Figures S1 and S2 might also show quite different spatial trends if TOC values are accounted for.

Thank you for the recommendation. We had originally opted not to normalize biomarker concentrations against TOC as not all samples were analyzed for bulk geochemistry due to insufficient sample, and for the ones that had been, it did not significantly alter the results. By focusing on concentrations, we could have a slightly larger dataset to work with. In any case, we do see the value of presenting this data, and plan to include it in the supplementary files and incorporate it into the discussion of the main text.

- Recommendations for paleoenvironmental reconstructions

This study highlights the complexity of biomarker signals in the highly dynamic Baffin Bay region, and our limited knowledge of their mechanistic behavior and applicability. Given my previous comments, and the uncertainty linked to comparing NOW vs. non-NOW sites based on concentrations of biomarkers per volume of sediment without accounting for sediment composition and in the absence of any form of age control, I cannot agree with the recommendation of proposing one type of biomarker (sterols) as a “more appropriate tool” (rather than HBIs) to characterize the NOW in the recent past. On the contrary, I think this study is a perfect example of why we need to pursue a multiproxy approach and not rely on single proxy lines of evidence. I would like to mention here that previous Holocene records from the NOW have mostly followed a multiproxy approach including microfossils, biomarkers, and biogeochemical proxies and I would be concerned if the community, based on this study, would go ahead and attempt to reconstruct the NOW based on sterols alone.

We appreciate this comment and the reviewer’s concern and will clarify in the text that a multi proxy approach is always recommended. Given our statistical analyses, the sterols indeed have significantly different concentrations in the NOW versus sites outside the NOW. This remains true whether they are presented as concentrations or normalized to TOC. While we agree that the analysis of other supporting proxies is needed to help elucidate why and how the NOW formed in the past (e.g., sea ice and temperature), as they are indeed responsible for its complex development, our results suggest that sterols will be useful to pinpoint when the NOW developed as it is today.

Detailed comments:

Lines 7 and 38 - Greenlandic Inuit is not a language. The correct term is West Greenlandic or Kalaallisut.

Thank you for highlighting this. We realized this ignorant mistake after submission, which has been now corrected. We also now include the local Eastern Canadian Inuit name, to be more inclusive and acknowledge all the traditional inhabitants of this paper’s study area.

Lines 33,34 – The instability of the NOW (e.g. Ribeiro et al. 2021) has been shown by a combination of multiple proxies, including lipid biomarkers, microfossils and bulk biogeochemistry (not just lipid biomarkers).

Absolutely, amended in the text.

Lines 44-45 – human occupation timelines are incomplete and outdated, please correct. See:

- Ribeiro et al. 2021 Fig 5 (already cited in this manuscript) and references therein, mainly: 1) Raghavan, M. et al. The genetic prehistory of the New World Arctic. Science **345**, 1255832

(2014). And **2)** Grønnow, B. & Sørensen, M. Palaeo-Eskimo migrations into Greenland: The Canadian Connection. In Dynamics of Northern Societies. Proceedings of the SILA/NABO Conference on Arctic and North Atlantic Archaeology, Copenhagen (eds Arneborg, J. & Grønnow, B.) 59–74 (National Museum, Studies in Archaeology & History, 2006).

Thank you for bringing this to our attention. The timing and references have now been updated accordingly.

Line 47 – This study should be mentioned here as well: Vincent, R. F. A study of the North Water Polynya ice arch using four decades of satellite data. Sci. Rep. **9**, 20278 (2019).

Done, and thank you for the suggestion.

Line 146 – specify coring device in the table per sample

We agree this is important information to include and have now added it to Table 1.

Line 149 – here, add any information on age control for the core tops if possible.

Unfortunately, we do not have any quantitative age control. We have added a clearer assumption of their recent ages to the Materials and Methods section 4.1.

Line 226 – did you mean to write "shoulder season months"? I am not familiar with this expression.

Yes, shoulder season refers to autumn and spring months. This has been clarified in the text.

Line 229 – TOC data should be added here. Specify in the table which of the 9 samples were analysed for bulk geochemistry.

Thank you for the suggestion, both will be added to the manuscript.

Line 307-308 – One could argue for the opposite, given that polynyas are characterized by intense sea ice formation, and the polynya area is under the influence of Arctic sea ice export. I suggest revising this section.

Apologies if we do not understand this comment fully, but we do describe these possibilities as suggested by the reviewer in the sentences immediately following L307-308.

Line 327 – Would be useful to add more information here on other potential sources for campesterol and b-sitosterol (besides marine diatoms). Also note that the Detleft et al 2021 study is in a fjord setting, while the datasets in this study excludes such settings. I don't completely follow the reasoning that correlation (of sterols) is supportive of a common source.

Apologies if our reasoning was not clear. We are aware that the Detlef study is from a fjord but wanted to mention it as it is the only other study that has reported campesterol and b-sitosterol from the greater Baffin Bay region. As we discuss in the Section 3 (Background on biomarkers) and in the sentences following L327 noted here by the reviewer, there are other potential sources of these sterols, such as vascular plants (Huang and Meinschein, 1976). While vascular *plants are commonly taken as the primary source for these sterols, we note that terrestrial biomass in this region of the Arctic is low (Gould et al., 2003), and without major river networks to transport the minimal biomass to the ocean, we assume that the contribution of these sterols from vascular plants is low compared to those produced by marine diatoms. Of course, this is only an assumption, but supported by the correlation of sterols in our surface sediment samples. Further detailed work on the production and transport of these sterols from algal and vascular plant samples in and around Baffin Bay will undoubtedly benefit paleoceanographic interpretations in future studies. We additionally mention this final point in the concluding paragraph on our suggestions for future studies.*

Line 350 – replace "all" with "partly" – biomarkers may partly originate from sea ice diatoms

Done.

Lines 365-366 – please revise this conclusion. Sterols alone are very unlikely to help us characterize the presence/absence of the NOW in the recent geological past, and other tools are available, such as "true" open water indicators.

Similar to the earlier comment by the reviewer, we will edit the text to clarify that multi proxy approaches are always favored to better assess the complex variables that lead to the NOW's past development. However, based on our statistical analyses (t-tests), sterols are the only biomarkers that are significantly different between the NOW and sites outside the NOW. While the other biomarkers in our study, as well as other proxies such as microfossil assemblages, provide important context for changes in the local paleoceanography, we believe based on our data that sterols will indeed be key to helping reconstruct the presence of the NOW.

Lines 472-473 – It is important to verify if this holds when accounting for TOC contents.

We note that importance of accounting for TOC as previously suggested by the reviewer, but the correlations between individual HBIs and sterols from the same sample would not change whether we normalize biomarker concentrations to sample mass or TOC. Therefore, this still holds true.

Line 479 – I suggest some caution here since sterols are not unequivocal open water indicators.

As we discussed in our previous response, our data and statistical analyses support this line of reasoning. While it is indeed true that sterols are not unequivocal open water indicators, our data suggest that sterols here may be dominantly sourced from open water primary producers in the area, rather than from other sources such as sea ice diatoms or vascular plants. This aspect will be further clarified in the main text.

Figure 4 – I suggest replotting with TOC normalized values

Due to the smaller number of samples that contain TOC information (n=9), we have added these suggested plots and comparisons with Kolling et al. (2020) to the supplement, rather than include in the main text.

Figure 5 – It took me a while to make sense of this figure. I suggest ordering the biomarkers in the same way as Fig 4 so they are easily comparable. IP25, HBI II, HBI III, HBI IV, Dinosterol,

Brassicasterol, Campesterol, b-sitosterol. Also specify if the plotted data in b) are all Kolling et al or a sub-set as indicated in the text? Sample sizes (n=) should be given for all figures.

Thank you for the good suggestions. Both will be amended accordingly.