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.

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.

- 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.

- 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.

- 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.

- 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.

Detailed comments:

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

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).

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,


Line 146 – specify coring device in the table per sample

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

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

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

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.

Line 227 – 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.

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

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

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

Figure 4 – I suggest replotting with TOC normalized values

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.