Reviewer 2 (Ruediger Stein):

The very well paper by Harning et al. is dealing with specific biomarkers determined in surface sediments from northern Baffin Bay, with a special focus on the North Water Polynya. These biomarkers, i.e., highly-branched isoprenoids (HBIs), sterols, long-chain alkenones and archaeal GDGTs, may give information about sea-ice conditions, open-water productivity, ocean temperature, and terrigenous input, as shown in many previous studies from very different ocean regions. Although major progress in using these proxies for reconstruction of present and past environmental conditions has been obtained, more ground truth data are still needed to fully approve some of these proxies, especially the proxies dealing with the reconstruction of sea ice (for example see reviews by Stein et al., 2012; Belt and Müller, 2013; Belt, 2018). Gaps in knowledge are related to the definite identification of the source of specific biomarkers, (regional) proxy calibration to allow (semi)quantitative estimates of sea-ice extent and its seasonal variability, sea-surface temperature, salinity etc. The missing ground truth data can be obtained by detailed studies of sediment trap material and surface sediments as well as cultural experiments and are especially needed from the high (polar) latitudes, e.g., the Arctic Ocean and its marginal seas.

In this context, the new data set from Harning et al. may give some important insight for using these biomarkers for characterizing the modern environmental conditions in a large polynya setting, i.e., an area of reduced sea ice and increased primary productivity. Such data might be strongly relevant for using the proxies for reconstruction of the development of past polynyas. Furthermore, Harning's data set allows to directly correlate a large variety of biomarkers (often either HBI and sterol data are produced/published from the same set of samples or alkenones or GDGTs). Despite this positive aspect, however, I have major problems with the present version of the manuscript. These concerns are related to the data set itself, the postulation of new SST calibrations for high latitudes based on a very limited data set, the missing more detailed comparison/discussion of the own data with the published biomarker data from Kolling et al. (2020), and some statements related to the interpretation of the biomarker data in terms of their origin (i.e., marine vs. terrigenous), as outlined in the following paragraphs. In my mind, a major clarification and revision of the manuscript is needed before it can be accepted for publication.

> We kindly thank Ruediger Stein (Reviewer 2) for his detailed review of our manuscript and highlighting some key areas for improvement. Below we provide responses to each of his comments and suggestions and look forward to submitting a revised and stronger manuscript.

Data base and presentation of data:

The total number of data points (13 samples in total) is quite limited for the quite general statements and interpretation of data given here. Whereas for the polynya itself (eight samples) this might be ok, it is quite questionable for the area outside the polynya (five samples for the entire northern Baffin Bay). Furthermore, these five samples are from very different settings (one from the central Baffin Bay, close to Davis Strait, and three off different fjord mouths), i.e., areas with very different sea-ice conditions and sedimentation rates. Due to the latter, these 1 cm thick samples (proxy data) might represent quite different time intervals (average sea-ice condition, temperatures etc. over 10 to hundreds of years). Furthermore, a presentation of the data in terms of distribution maps (see Kolling et al., 2020) might be useful here as well. In order

to allow more general statements/interpretations/conclusions, the discussion of the new data together with the data from Kolling et al. (2020) would be most helpful/important. Kolling et al. (2020) have studied samples from locations very close to those of this study (Fig. R1). Thus, data from the same biomarkers (i.e., IP25, HBIs and sterols as well as the different PIP25) are available for a detailed comparison and interpretation in terms of sea-ice extent and its seasonality, productivity etc. (see examples in Fig. R1).

While we are fully aware of the limited size of our dataset, it was what was made available for our study, especially in the polynya region where we plan to apply these calibrations for paleoceanographic reconstructions. The reviewer is certainly correct that the different locations and environments of the samples can lead to different integrated intervals of time, which we currently highlight in L159 of the manuscript. However, this is an issue for any surface sediment proxy calibration that covers regions with differing sedimentation rates and will always introduce added uncertainty. By acknowledging this limitation as we currently have, we hope this satisfies the reviewer's concern.

We do currently provide distribution maps for the HBIs and sterols (please see the supplemental material), however, we prefer not to interpolate the data as done in Kolling et al. (2020) to avoid over-interpretations related to spatial gaps in the datasets. Following the additional analytical steps outlined in our response to the following comment, we indeed plan to make more direct and integrated comparisons with the Kolling et al. (2020) datasets. We appreciate the opportunity to expand on this in the revised manuscript.

Low concentrations of HBIs and sterols:

When I myself tried to roughly compare the new Harning et al. data with our Kolling et al. (2020) data, I realized a "problem", and this is one of my major concerns I have with the new data presented here. The absolute concentrations of the HBIs and the sterols are significantly lower (two orders of magnitude or so!!) than those presented by Kolling et al. (2020). What might be the cause for this? Storage of samples (fresh/deep-frozen samples vs. samples stored under room temperature), or different analytical approach? I am not a chemist. Thus, I myself cannot comment in detail the analytical approach for identification and quantification of the biomarkers that has been used here, but I know from cooperation with the chemists involved in our biomarker analyses as well as the Simon Belt group that the analytical procedure is overall significance if data from different labs will be compared. Thus, Belt et al. (2014) carried-out an inter-laboratory investigation dealing with the identification and quantification of the Arctic sea ice biomarker proxy IP25 and other HBIs in marine sediments (our lab was involved in this study as well). As final statements these authors summarized in their abstract that "data are presented that suggest that extraction of IP25 is consistent between Automated Solvent Extraction (ASE) and sonication methods and that IP25 concentrations based on 7-hexylnonadecane as an internal standard are comparable using these methods. Recoveries of some more unsaturated HBIs and the internal standard 9-octylheptadecene, however, were lower with the ASE procedure, possibly due to partial degradation of these more reactive chemicals as a result of higher temperatures employed with this method. For future measurements, we recommend the use of reference sediment material with known concentration(s) of IP25 for determining and routinely monitoring instrumental response factors."

Harning et al. have used the ASE for extraction, and they should check their HBI and sterol data and comment on their analytical approach. For example, did you take in account the different response factors for the analytes and internal standards?

We thank the reviewer for this comment, and for the details of previous work on extraction techniques and quantification biases. The reviewer is indeed correct that the difference in analytical approach and storage of samples can both lead to underestimation of biomarker concentrations (Belt et al., 2012; Cabedo-Sanz et al., 2016). We used a methylated alkane (3-methylheneicosane) that our lab normally uses for aliphatic hydrocarbons. Unfortunately, we were not fully aware of how different the response factor of this standard could be compared to the HBI standards used by Belt et al. (2012) and Kolling et al. (2020). We agree that the structural differences will lead to different ionization efficiencies and response factors, which can alter the calculation of HBI concentrations. In order to make our HBI datasets comparable to those of Kolling et al. (2020), we propose several steps that we will complete during the revision period:

1) We have contacted the Belt lab to procure the 7-HND standard and will run a 7-HND dilution series under the same GC-MS operating conditions as we did for our samples to calculate its response factor. We will compare this against that of 3-methylheneicosane.
2) The Belt lab has also provided us with sediment that contains a known concentration of the different HBIs that we will analyze on our GC-MS. We aim to quantify HBIs in this sample using response factors from 7-HND and 3-methylheneicosane,
3) By comparing the differences in response factors (step 1) with the peak areas identified by their molecular ions in SIM mode (step 2), we can correct our initial response factors to produce corrected concentrations.

While the standards used for sterol quantification are less standardized, we do note that Kolling et al. (2020) used a different standard for quantification in their study (cholesterol-D_6) than the one we used (1-nonadecanol). Therefore, to make our datasets more comparable, we will follow a similar approach as outlined above for HBIs. More specifically, we will generate dilution series for several sterols (cholesterol, ergosterol and stigmasterol). The cholesterol standard will allow us to correct our brassicasterol and dinosterol concentrations, both of which were analyzed by Kolling et al. (2020). The dilution series for ergosterol and stigmasterol (two phytosterols) will allow us to test if the ionization efficiencies in these structurally similar sterols is akin to cholesterol's. If they are indeed similar as expected, then we can adjust the concentrations for our campesterol and ß-sitosterol accordingly. However, Kolling et al. (2020) did not analyze these latter phytosterols, so no comparison can be made.

Finally, if our proposed approach to correct our HBI and sterol concentrations does not bring the two different studies into alignment, then we will have to consider sample storage and/or extraction procedure as additional possible factors. In this regard, we note that Kolling et al. (2020) stored their sediment samples in glass vials or plastic bags below -20°C. On the other hand, our sediment samples were stored in glass vials but at 4°C. Previous studies have noted that biomarkers, such as HBIs, can degrade faster at warmer temperatures, although this was more readily observed at room temperature (Cabedo-Sanz et al., 2016), and thus, may not be a concern in our study. The extraction method (ASE vs ultrasonication) may also result in preferential degradation of lipids but would not be able to be corrected for. If such is the case, the relative abundance of the HBIs should be comparable to those published Kolling et al. (2020), and comparison between the two datasets could be achieved accordingly.

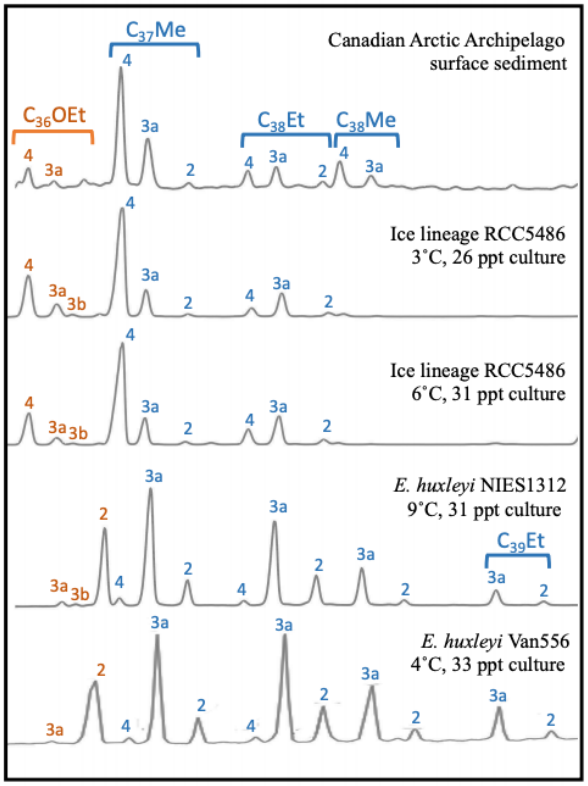
Alkenones, high UK37', and new SST calibration:

Concerning the alkenone data, I also have problems. The northern Baffin Bay is an environment with sea-surface temperatures significantly lower than 10°C (<5°C). Under such cold conditions,

the C37:3 long-chain alkenones should be predominant over the C37-2 alkenones (e.g., Prahl and Wakeham, 1987). In this study, however, the C37:2 alkenones are predominant (Fig. 4 of their paper), resulting in high UK37' values of 0.7-0.9 (Fig. 7 of their paper), i.e., values that are typical for much higher temperatures. In polar and subpolar regions including Baffin Bay (!), however, UK37' values are typically between 0.2 and 0.4 (e.g., Rosell-Melé, 1998; Bendle and Rosell-Melé, 2004; Méheust et al., 2013; Moros et al., 2016). The authors are aware about this problem and discuss that temperature, salinity and nitrate are factors that may have influenced UK37' and UK37 values. Nevertheless, they state that "given our limited dataset at this time, temperature seems to be the most important environment variable on UK'37 values", and then simply have correlated the high UK37' values with the measured low SST values of Baffin Bay. As result, they obtain, of course and not surprising, a new calibration that give totally different SST values in comparison to those calculated based on Müller et al. (1998) (Fig. 7). Did the authors test their "new calibration" using the UK37' data from Moros et al. (2016), i.e., data from Baffin Bay? What SST values they would get? Furthermore, there might be another option, i.e., there might be some question mark related the data that should be clarified before postulating a new calibration (e.g., what about co-elution with other compounds that might result in too high C37:2 concentrations; see Villanueva and Grimmalt; 1996). Please check!

The reviewer brings up an important point of discussion here about the distribution of alkenones in our dataset. While their distribution, particularly the high abundance of $C_{37:2}$, in our dataset was certainly surprising, there is considerably less known about alkenones at higher latitudes compared to mid and low latitudes. Without additional data from northern Baffin Bay to compare with, we hypothesized that other environmental factors (e.g., salinity, nutrients) could be influencing the patterns we observed. We were also hesitant to test the calibration on the Moros et al. (2016) dataset as the focus area of our calibration in northern Baffin Bay is distinct in terms of oceanography from Moros' of west Greenland. In northern Baffin Bay, multiple water masses (Polar, Arctic, and Atlantic) converge whereas west Greenland is only under the influence of the predominately Atlantic-derived West Greenland Current, which could lead to distinct alkenone distribution patterns and make the application of a northern Baffin Bay calibration inappropriate.

That being said, we were unaware of some surface sediment samples from the Canadian Arctic Archipelago that were recently analyzed for alkenone distributions (Wang et al., 2021). These distributions are shown in the figure below (top panel) obtained from the supplementary material in Wang et al. (2021). Given that these distributions are consistent with the accepted relationship between alkenone unsaturation and temperature, and not with the distribution observed in our dataset, we will need to revisit our analyses. In particular, we plan to check for potential co-elution with $C_{37:2}$ as suggested by the reviewer.



In the introduction (Lines 55-57), some credit should be given to other studies dealing with the use and calibration of UK37 and TEX86 for SST reconstructions in high latitudes (Sikes et al., 1997; Rosell-Melé, 1998; Ho et al., 2014)

We agree with the reviewer that citing these previous studies is important and will be included in the revised manuscript.

Biomarkers (sterols) and their sources:

I do not agree with the general statement that in the Arctic the abundance of brassicasterol, dinosterol, campesterol and ß-sitosterol is mainly related to marine productivity. The interpretation of the sterols and their use as organic-carbon source indicator are not easy tasks and may strongly differ from region to region (e.g., Volkman, 1986; Huang and Meinschein, 1976; Fahl and Stein, 1997, 1999; Belt et al., 2013). Brassicasterol often used as “marine productivity indicator” might be ok in areas not influenced by strong terrigenous input (river discharge). In coastal areas controlled by huge river discharge, such as the Kara and Laptev

seas, a large amount of brassicasterol found in surface sediments have a terrestrial (lacustrine) source (e.g., Fahl et al., 2003; Hörner et al., 2016). Furthermore, in these shallow-water coastal zones, terrestrial/lacustrine brassicasterol as well as brassicasterol produced by marine algae may be incorporated into sea ice and transported into the open central Arctic Ocean. Thus, these biomarkers not produced by sea ice may be found in the sea ice and result in erroneous interpretation of the source of this biomarker. When using brassicasterol as "open-water productivity proxy", additional information about the environmental situation should be taken into account and additional biomarkers (such as dinosterol and the HBI-III) should be used as well. When using the PIP25 approach for reconstructing sea-ice conditions, PIP25 values based on IP25 and brassicasterol, dinosterol and HBI-III, should be calculated and discussed.

Vascular plants are producers of campesterol and ß-sitosterol (Huang and Meinschein, 1976) but may also be produced by diatoms (Belt et al., 2013). In the Arctic Ocean characterized by strong terrigenous input into the marginal sea and – via sea ice and ocean currents – a predominantly terrigenous source of these biomarkers is most probable (e.g., Yunker et al., 1995, 2005; Stein and Macdonald, 2004; Xiao et al., 2013). Such riverine input of organic carbon onto the shelf is nicely reflected in maximum concentration of campesterol and ß-sitosterol in surface sediments close to the major river mouths in the Kara and Laptev seas (Xiao et al., 2013; Fig. R2). In most part of the Arctic Ocean, thus terrigenous organic matter is predominant in surface sediments (Stein and Macdonald, 2004; Fig. R2). Thus, I. cannot agree with the authors' statement in 113/114): "In the Arctic where terrestrial biomass is low, we assume that the contribution of terrestrial-derived campesterol and β-sitosterol is minimal compared to that produced in the ocean."

The reviewer brought up an important concern in our statements of sterol origins, which we realize is likely due to a lack a specificity on our part. In other regions of the Arctic, such as the Kara and Laptev Seas, we are aware that there can indeed be a significant terrigenous component of sterols to marine sediment, particularly in the form of campesterol and ß-sitosterol. However, we believe the confusion here relates to the fact that we initially referred to the "Arctic" in general. In the revised manuscript, we will be more specific and state in L113 that "In the Canadian Arctic Archipelago where terrestrial biomass is low (Gould et al., 2003), we assume that the contribution of terrestrial-derived campesterol and ß-sitosterol is minimal compared to that produced in the ocean". In addition to the low terrestrial biomass, there are also no rivers that drain into Baffin Bay that are comparable to those that drain the Siberian Arctic for instance. Hence, the combination of the Canadian Archipelago's low terrestrial biomass and river network, the fact that all sterols are strongly correlated (Fig. 5), and the overall marine signature of the sediment as indicated by bulk geochemistry (Fig. 3) all strongly suggest that at least in northern Baffin Bay, the 4 sterols we analyzed are all dominantly synthesized by marine organisms.

References:

Belt, S.T., Brown, T.A., Navarro Rodrigues, A., Cabedo Sanz, P., Tonkin, A., and Ingle, R.: A reproducible method for the extraction, identification and quantification of the Arctic sea ice proxy IP_{25} from marine sediments: Anal. Meth., 4, 705-713, 2012.

Cabedo Sanz, P., Smik, L., and Belt, S.T.: On the stability of various highly branched isoprenoids (HBI) lipids in stored sediments and sediment extracts: Org. Geochem., 97, 74-77, 2016.

Gould, W.A., Raynolds, M., and Walker, D.A.: Vegetation, plant biomass, and net primary productivity patterns in the Canadian Arctic: J. Geophys. Res. Atmos., 108, 1-14, 2003.

Wang, K. J., Huang, Y., Majaneva, M., Belt, S. T., Liao, S., Novak, J., Kartzinel, T. R., Herbert, T. D., Richter, N., and Cabedo-Sanz, P.: Group 2i Isochrysidales produce characteristic alkenones reflecting sea ice distribution. Nat. Comm., 20, 1-10, 2021.