

Interesting data are presented, however, the manuscript needs major revision. It is not clear what the data actually tell us, and why. One major issue is that it is not clear what the authors actually want to present. The title suggests that the main aim of the paper is to present a new age model and the introduction is mainly written with a focus on the need for a new age model. Secondly, the title hints at an asynchronous relation between marine and terrestrial climate. Even though both marine and terrestrial data is presented I do not see a focus on an asynchronous relation reflected in the paper. The authors need to figure out if the aim is to present an age model paper or a climate paper, and structure the paper accordingly.

A: We thank the reviewer for their compliments on our dataset, and constructive comments on the manuscript. We appreciate their concern about what the data is telling us and why, and we add that this is quite a challenging dataset to interpret. The main focus of the paper, as you have determined from the first part of our title, is to present an age model for the Pliocene of the North Sea Basin. We have identified the need to construct an age model, which is missing from this region for the Pliocene, in order to place a variety of new and existing proxy records into context (e.g. L113-117). We agree that the discussion based on the asynchronous shifts of marine and terrestrial climate is not at the focus of the paper, and comes as part of a more speculative section near the end in section 5. Therefore, based on your comment and that of reviewer #1, we will amend the title of the paper to: "A new age model for the Pliocene of the Southern North Sea Basin: a multi proxy climate reconstruction", and adjust the focus of the discussion and conclusions accordingly.

The discussion needs to be transferred from providing statements to become a discussion. What do your results show? How can they be explained/what can they tell? Why? What is the suggested mechanism? What is the argumentation for why that possible mechanism is most likely – and for others being less likely? When stating relations to other records, why don't you show some those records as well so that it can be evaluated?

A: We are a bit confused by this comment, as we feel that we are doing this in sections 4 and 5. Section 4.1 discusses the logic and reasoning behind the age control that we present in Fig. 5, and involves the results from the palynology records, the seismic profile, the foraminifera isotopes, and the gamma ray logs ('What do the results show?', and 'How can they be explained?', and 'What is the mechanism?'). This discussion subsequently leads us to identifying the best interval suited for tuning to the LR04 benthic stack. Subsequently, section 4.2 presents a discussion on the potential confounding factors on the sea surface temperature (SST) records generated by the different lipid biomarkers, how the records should be interpreted, and why (i.e. explanation, comparison, mechanisms). We then use section 5 to discuss and interpret our multi-proxy climate records in the constrained age domain, and to place them in a global context (comparison and mechanisms). With regards to showing other records mentioned in the discussion section, see the specific reply to the comment below. Nevertheless, we will carefully read the discussion again and further clarify our lines of reasoning where possible in the revised manuscript.

The figure that presents the basis for the new age control is impossible to evaluate and needs to be redone. I would prefer to have the datasets correlated to each other shown underneath each other in the same direction so that the relations between the ups and downs of the two records are immediately visible. I would also like to see how the two records (LR04 and yours) match after correlation, showing both records towards age.

A: We feel like the gamma ray log and lithostratigraphic log on the figure are good additions for the reader to place the oxygen isotope record of Hank into context, since the foraminiferal isotopic composition in such a shallow marine succession is dependent on other factors, such as depositional environment and hiatuses (see discussion in section 4.1). The lithostratigraphy and the gamma ray log can both give insight into these important processes (e.g. sharp inflections on the gamma ray log may indicate a hiatus). Similarly, the backdrop of the biostratigraphic age tying points is also useful to evaluate the oxygen isotope-based age model with the initial one presented in Dearing Crampton-

Flood et al. (2018). Therefore, we feel that Fig. 5 warrants the addition of more information than just the age-tying points and the two oxygen isotope records. That being said, our revised evaluation of the periods where we suspect that ice volume changes may not play a big role (see discussion response to 'Do you use the d13C data at all?' below) will cause us to re-evaluate the tuning that we present in Fig. 5.

We will add a figure with the LR04 stack and the Hank oxygen isotope record on the same x-axis (age) to the supplementary information (Fig. S2) of our revised manuscript.

Some of the figures present a lot of data in a small format. Check font sizes and make sure they are readable.

A: We apologize if the labels in the figures are difficult to read. We will assess and optimize the font size of all figures where appropriate (Figs. 3 and 5 in particular).

Do you use the d13C data at all?

A: The $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ records give an indication about possible ice volume influences, given that the trends in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ mirror each other (deep ocean concept). This is exemplified in the record of Noorbergen et al. (2015), who observed mirroring trends in the shallow marine Quaternary sequence from Noordwijk. When the trends between the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ records do not mirror each other, the ice volume signal is less dominant due to other factors which overrule the signal such as freshwater influence, reworking, and diagenetic influences (the latter assumed to be minimal because only well-preserved foraminifera were picked). Therefore, including the $\delta^{13}\text{C}$ record and comparing to the $\delta^{18}\text{O}$ record will aid in demonstrating where ice volume signal is recorded in the Hank record and when it is not. We will include a sentence of where the two isotope records mirror each other in the results section (section 3.1) of the revised manuscript. Further, we will include a summary of this response into the discussion section of the paper in order to improve the evidence/statements/discussion that lead to the targeting of the tuned interval and therefore the creation of the final age model.

Comments that emerged while reading the manuscript:

L23:costal zones, linking the costal and continental climate evolution, are lacking.

A: Thank you for the suggestion, we will make the suggested change.

L30: ... stable oxygen isotopes...

A: We will make the suggested change.

L57: Fewer and less well constrained temperature records exist for the terrestrial realm (Zagwijn, 1992; Salzmann et al., 2013), however, they all indicate that climate was warmer than present. Why do you consider them to be less well constrained? Should be specified – because of age control? Because of other issues?

A: The temperature records presented by Zagwijn (1992) and compiled by Salzmann et al. (2013) are derived from pollen assemblages. Pollen based temperature estimates are generally less constrained than that of e.g. SST reconstructions, as i) they present a temperature range rather than an absolute temperature estimate (Coope, 1970, Ann. Rev. Ent.; Mosbrugger and Utescher, 1997, Palaeoeco. Palaeoclima., Palaeoeco.), ii) they can be skewed toward growing season (Guiot, 1990, Palaeoeco. Palaeoclima., Palaeoeco.) and, iii) if based on terrestrial sediment sequences, can suffer from poor

age control, as terrestrial outcrops are more difficult to date than marine sediment sequences. We will add this brief explanation to the appropriate section (introduction) in the revised manuscript.

L106: I would be more careful than to say that the infaunal species are unaffected. They can trace large scale climate inflicted changes equally well as epifaunal species, e.g. detect brine signals during glacial times (e.g. Dokken et al., 2013). Furthermore, you discuss a potential fresh water influence later on, so some inconsistency here.

A: We thank the reviewer for noticing this inconsistency. We will modify the sentence in the revised manuscript to read: "The depth habitat of endobenthic foraminifera in the sediment provides a moderate degree of shelter from disturbances such as reworking by bottom currents and freshwater input". This revised sentence provides a more accurate view of the role that the endobenthic foraminifera play in climate reconstructions.

L118: What is relative land cover?

A: We use the term relative land cover to describe the prevailing vegetation cover on the nearby continent during the time of deposition. We change this into 'prevailing vegetation' in the revised version of our manuscript.

L134: Your site is very close to the coast and I would expect much lower water depths than in the central basin. Do you have any idea about what the water depth was at your site? In such a shallow setting the water depth have implications for how to interpret your records.

A: According to the height of the clinoform as observed in the seismic profile (Fig. 4), the water depth in the interval ~300-200 m is approximately 80–100 m. The clinoform stacking suggests a continuous westward progradation of the coastal system, such that this location became progressively infilled by sediment. Note that this estimate does not refer to the downdrop which we associate with the (possible) M2 event or recovery thereof. We will add a few words to the end of the sentence to remind the reader that the Hank site in the Southern North Sea Basin is expected to have water depths lower than that predicted for the central basin. Later in the results and the discussion section where the seismic section is introduced and described we will specify the estimated water depth for the Upper Oosterhout formation (80–100 m).

L145: Why is it an advantage that you get a smoothed record?

A: The advantage of this method of core collection is that the 1 m resolution leads to smoothed records compared to the relatively expansive total record (404 m for the Hank Core). The possible disturbances generated from the dynamic environmental setting of the Hank site at this time will be smoothed, which makes it easier to reveal the more regional climate signal (albeit in lower resolution).

L177: Did you rerun any of the samples you consider to be outliers? What is the argumentation for this choice for removing data point? It seems a bit arbitrary. What will the record look like if you include all measurements? What seems to be outliers can very well be true values if it just looks like the values don't belong.

A: We realize that since the $\delta^{18}\text{O}$ values of foraminifera recorded at the Hank site may not necessarily be attributed to ice volume signal alone, a different approach should be taken to identify outliers, given that freshwater input and reworking may affect the $\delta^{18}\text{O}$ signal (see reply to earlier comment). Unfortunately, we did not rerun the samples we considered to be outliers. This hampers the interpretation of the record, especially in light of the extremely high and low $\delta^{18}\text{O}$ values recorded

(grey triangles in Fig. 2). Thus, we will remove the method of identification of outliers from section 2.2 in the manuscript, and update section 3.1 with revised values (ranges etc.) including the whole record. This will also involve replotting the $\delta^{18}\text{O}$ data in Fig. 2, and Fig. 5 by extension.

The tuning for the age model is not expected to change largely because the majority of the tuned interval (296–206 m) does not contain that many outliers compared to the rest of the record (grey vs. black data points in Fig. 2). The tie points that are used particularly in the depth interval ~270–240 m will not change, particularly as this is the depth interval where a mirroring trend is seen between the $\delta^{18}\text{O}$ and the $\delta^{13}\text{C}$ records (see earlier response).

L250: What about C_{37:4} as a fresh water indicator? Do you see similar changes there? I see that you state later that no C_{37:4} is present. Implications? Why do you see changes in one of your freshwater indicators but not in all?

A: As we mentioned in L558-561 we did not detect the presence of the C_{37:4} alkenone in our samples. The absence of C_{37:4} in the mid-Pliocene (300–200 m) part of the interval does seem to fit with the other environmental indicators such as (i) low terrestrial/marine (T/M ratios), (ii) low BIT index values, and (iii) the presence of marine biomarkers such as Crenarchaeol and long chain diols (Figs. 2, 3). However, it is perhaps surprising that no C_{37:4} was detected in the interval '3' (Fig. 2), which corresponds to the late Pliocene/early Pleistocene, as the palynological data indicate a high T/M ratio and the presence of freshwater and brackish water algae species (L310-313). However, transect studies from surface sediments in the Baltic Sea indicate that the percentage abundance of the tetraunsaturated C_{37:4} alkenone only appears higher (>15 %) at salinities around 8 psu and below (Bornholm Basin; Schulz et al., 2000; GCA; Kaiser et al., Org. Geochem.). Therefore, we interpret that the salinity change that affected the Hank Site as a result of a gradual change to a more estuarine environment did not change the salinity significantly enough to lower it below 8 psu. On the other hand, a very speculative reason behind this could be the affinity for the alkenone-producing organisms to the type of estuarine environment that the Hank site progressively became over the Plio-Pleistocene transition. However, until more is known about the specific organisms that produce these lipids, not much more can be said at this time.

L255: extracted and processed for what?

A: We will change this sentence into: 'lipid biomarkers were previously extracted from the sediments (n=155) and separated into polarity fractions according to ...'

L283: What/why was it challenging?

A: The low abundance of foraminifera in the crag material made it challenging to pick in that interval. We will add this explanation to the revised manuscript.

L290: Here you state a variability of approximately 1‰ while based on the information above (L285 and onwards) its closer to 2.6‰ (or 0.9 to 1.8‰ if you selectively look at specific transitions). And why would you expect the d13c variability (in amplitude) to match the d18O variability?

A: We apologize for the confusion and will remove the reference to $\delta^{18}\text{O}$ variability in L292-293, as the estimate in L286-287 is the more accurate representation of $\delta^{18}\text{O}$ variability in this record. However, in light of including all the $\delta^{18}\text{O}$ values measured (as discussed above), the reported variability in the manuscript will change. Secondly, as far as we are concerned, we did not phrase the sentence to suggest that we expect the amplitude of $\delta^{13}\text{C}$ variability to match the $\delta^{18}\text{O}$ variability. We will rephrase this sentence as: 'Discounting the sample at 206 m, the variability in the $\delta^{13}\text{C}_{\text{cass.}}$ record is approximately ~1 ‰ (Fig. 2d).'

Line 336: Lipid biomarkers and proxies. Why do you include “and proxies” here? The isotopes, palynology and biomarkers are all proxies.

A: We apologize for the confusion, the ‘proxies’ we refer to here refer to the specific proxies that are based on lipid biomarkers. We understand your confusion and will rename the section to read: ‘Lipid biomarkers’.

L347: This is a very large range. Given your choice regarding “outliers” for the oxygen isotope record, why do you keep the biomarker results as measured? Supports the feeling I get that the removal of some isotope measurements are done a bit randomly– even though you have a set way of defining which points you removed.

A: Thank you for the comment. We agree that the Uk37 range we have calculated throughout the core is very large. We indeed mention the large range of reconstructed SSTs based on Uk37 in L347-353, L564-575, and L651-671, where we also compare the data with other SST records based on Uk37 in the North Atlantic, which also show a large range in SSTs. This is in contrast to the $\delta^{18}\text{O}$ data, for which the range of variation is not reported elsewhere. We feel it is appropriate to plot the samples considered ‘outliers’ in the record so the reader can judge for themselves. This is why we have initially plotted those $\delta^{18}\text{O}$ outliers in grey in Fig. 2c in our discussion paper, and also why we will include all $\delta^{18}\text{O}$ measurements in our record in Fig. 2 and 5 in the revised manuscript.

L355: selected

A: We will not follow this suggestion as we believe this to be the correct usage of the term ‘select’ in this context.

L374: Why do you get a stronger terrestrial influence towards the Pleistocene? Increased glacial erosion?

A: The stronger terrestrial influence towards the Pleistocene at this site is most likely a result of sea level drawdown caused by the increasing volume of ice build-up in the northern hemisphere. The Hank site is a shallow marine succession which borders on the area between river/delta deposits and marine sands (Fig. 1 in manuscript), thus the progradation of the Rhine-Meuse River during this time would have brought more terrestrial material to the Hank site and contributed to the changing depositional environment during the transition into the cooler Pleistocene.

Line 411: MAT. Is this short for mean annual temperature? Not mentioned before, needs to be defined. Since you refer to your 1018 paper I assume the record is published and the method is described there, however, I am still curious about how certain you are regarding the absolute values presented given that your record includes extinct species, e.g. *Sciatopitys*? You refer to several other terrestrial profiles – can these be shown for comparison? If they lack age control, can you really link them to your record/state that it is the same?

*A: MAT is indeed short for mean annual temperature, which we will specify in our revised manuscript. The MAT record has been published in Dearing Crampton-Flood (2018) and is based on temperature-sensitive membrane lipids of soil bacteria (branched GDGTs), which, in principle, yield absolute temperature estimates. A thorough discussion on the reliability of the absolute MAT values can be found in the same paper. Note that we have not reconstructed temperature using pollen distributions. We also note that species *Sciatopitys* is considered a relict taxon as it is not yet extinct (occurs in Japan). For a discussion regarding the age correlation between terrestrial profiles, see the discussion in the author responses to reviewer #1.*

L438: M2 is very pronounced in some records, but there are also several sites where its less pronounced, making it an enigmatic period with large uncertainties related to the magnitude of the “glacial” event. Risebrobakken et al., 2016 consider the possibility of a hiatus as an explanation for the lack of the most extreme signature but also that it might have been a less extreme event than expected from e.g. LR04.

A: We have indeed considered whether the M2 event is a less extreme event than in some of the other areas of the world. However, there are many papers dealing with the stratigraphy of the North Sea that suggest a hiatus during MIS M2, particularly in the southern North Sea Basin (De Schepper et al., 2009, Geol. Mag.; Louwye et al., 2010, Geol. Mag.). See also the comment that Stijn de Schepper posted in the discussion. We also consider the occurrence of a hiatus during M2 at the Coralline Crag formation which is close to the Hank site (Fig. 1, manuscript), to explain the sequence boundary in the seismic profile at Hank. The position of this boundary indicates that the hiatus covers most of the M2 at the Hank site. We will expand the discussion in the revised manuscript to include the evidence for a hiatus over the M2 event in the adjacent sites in England and Belgium to support our interpretation. We will also add that Risebrobakken et al. (2016) considered that the M2 may have been a less extreme event in this region compared to other regions in the world where it is more pronounced, to provide the reader with a more nuanced view on the evidence for these different hypotheses on the M2 event in this region.

Line 443: I cannot see that the North Sea is mentioned specifically in Miller et al., 2011? Overall, estimates of sea level change during M2 varies tremendously in literature. This should be acknowledged. Also take into account the findings of Raymo et al., 2018 where they conclude that for the Pliocene, geochemical sea level proxies currently carry uncertainties too large to allow any meaningful ice volume (hence sea level estimates).

A: We apologize for the confusion; we will remove the specific reference to the North Sea in this sentence. We agree that the estimates for sea level change in the literature that correspond to the M2 event vary greatly. We also thank the reviewer for directing our attention to Raymo et al. (2018), and we will modify the sentence to read: ‘There is evidence for a large global sea level drawdown (estimates of 70 m; Miller et al., 2011) during the M2, however large uncertainties in the estimation of ice volume prohibit any meaningful estimates of sea level for the Pliocene using the stable isotope measurements of foraminifera (Raymo et al., 2018).’ However, the estimation of water depth at the Hank site is based on the seismic profile, and is approximately 80–100 m (see reply to earlier comment). Other estimates indicate that Pliocene sea level in the North Sea Basin was approximately 60–100 m (Overeem et al., 2001; Basin Res.), which agrees with our interpretation. These estimates however do not provide any insight into the sea level drawdown before that, associated with the possible M2 event or recovery of (corresponding to the sequence boundary in Fig. 4). We will include all relevant discussion on this point and references into the revised manuscript.

LL461-472: You should not forget that your data is from a very shallow site and hence cannot be expected to reflect the same absolute values as the intermediate and deep water records from Risebrobakken and LR04. How do you physically transfer the suggested freshwater signal from the rivers to the bottom of the North Sea? A river signal will be much less dense than a marine water mass and lay as a fresh lid on top of the denser water mass. This is one example of where knowledge of the paleo depth of your site is key to be able to make trustworthy interpretations of your data.

*A: We appreciate the comment and we agree that the absolute $\delta^{18}\text{O}$ values cannot be expected to be comparable to those of other records from cores recovered from deeper depths in the open ocean. We also consider the study of Noorbergen et al. (2015), who constructed a Quaternary age model from a similarly shallow succession in the North Sea, and found that the absolute $\delta^{18}\text{O}$ values of a composite $\delta^{18}\text{O}$ record of *Bulimina aculeata*, *Cassidulina laevigata*, *Cibicides lobatulus*, and *Elphidiella hannai*, were comparable to the LR04 stack. Hence, we tried to adopt the same approach*

to anticipate the influence of freshwater input by only picking endobenthic foraminifera in the Hank sediments (cf. Noorbergen et al., 2015). Hank is located relatively closer to the mouth of the paleo Rhine, and thus is likely associated with a shallower water depth (see response to previous comment for paleodepth interpretation of 80–100 m), which we believe is reflected in the large range of $\delta^{18}\text{O}$ values, and thus likely also the offset compared to the LR04 stack. We will add the paleodepth interpretation to the revised manuscript. With regards to the transport of fresher water to lower depths in the North Sea, we note that the North Sea is a high energy system where wave action and winnowing contribute to the mixing of freshwater input in the relatively shallow water column (Charnock et al., *Understanding the North Sea System*, Springer Science and Business Media). We will include this explanation in the revised manuscript.

L479: The difference in amplitude of change between the global deep water stack and your shallow regional/coastal site does not prove that the area is more sensitive to climate disturbances. Furthermore, what do you mean by climate disturbances? This is an empty term. This section is also an example of statements without discussion. You should check this throughout and make sure that your discussion is a discussion and not just statements. Discuss your results, potential explanations, reasons for why one or the other potential explanation is more or less likely, and in the end conclude on what you find to be the most likely solution and why. The mechanisms are important. E.g. how do you physically make the fresh water reach the bottom of the North Sea in order to explain light benthic isotope values by fresh water influence.

A: *The larger amplitudes ($\Delta \delta^{18}\text{O} > 2.5 \text{‰}$) observed in the Hank record are not the result of a dominant single climate factor such as ice volume. This is in contrast to case in deep ocean settings. At the Hank site, freshwater influence is more likely to lead to the large observed amplitude in $\delta^{18}\text{O}$. We will therefore adjust all language referring to 'climate disturbances' in the revised manuscript, and include the interpretation of the large amplitude as a result of periods where the Hank site was particularly sensitive to freshwater input, probably deriving from the proto Rhine-Meuse system. A literature search shows that similar observed amplitudes and associated low benthic and planktic $\delta^{18}\text{O}$ values were recorded in marine sediments from the Ionian Sea, and were interpreted as freshwater influence that coincided with sapropel deposition (Schmiedl et al., 1998; *Paleoceanography*). We will inspect the discussion section carefully and include a more thorough discussion into this perceived freshwater input, with appropriate evidence and references. Further, the interpreted paleodepth (80–100 m) during the mPWP period (~300–200 m) will be added to the revised manuscript which leads to a better supported discussion on why the Hank site was sensitive to fresh water input, due to its shallow nature.*

L496: When you say that you tune the warmer periods, what exactly do you mean by that? I would never use the maxima or minima as tuning point between records, but rather go for transitions, since the character of the records you tune is bound to be different.

A: *We realise that we have not followed the method that is mostly used for age model constructions, however, the coastal location of the Hank site requires a slightly different approach, as is outlined in the manuscript (section 4.1). The combination of a relatively weak influence of the Atlantic and the potential occurrence of hiatuses during cooler phases, which could also include (part of the) transitions, has led us to use the warmer periods for tuning. Warmer periods in the record are identified as the maxima in the $\delta^{18}\text{O}$ record (Pliocene interglacials: G17, G15, K1 etc. on LR04 stack). Furthermore, see the discussion above on the sampling resolution (1 m) of the Hank borehole, which may preclude the capture of any true inflection point or transition in the $\delta^{18}\text{O}$ record.*

As discussed above, we will identify the intervals of the record with opposing or mirroring trends in $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ and only use those for tuning the warm intervals to. This will lead to changes the discussion section and Fig. 5. We will therefore include this discussion, and a revised and updated form of Figure 5 in the manuscript.

L505-513: From figure 5 it is impossible to evaluate the correlation between LR04 and your record and hence the basis for your age model. This figure needs to be improved. One suggestion would be to show the record so that you read them in the same way. Another thing I would require to see is a direct comparison the two isotope records vs age, following your new tuning of the record.

A: As promised in our reply to an earlier comment, we will include a figure in the supplement that shows the correlation between the LR04 and the $\delta^{18}\text{O}$ record from Hank, on (A) separate depth/age scales and on (B) the same age scale (Fig. S2).

L527-528: I would expect that a dominance of deeper dwelling organisms influencing the GDGT data will provide colder temperatures than what you will expect for the surface, even if the water depth is shallow. At what depth is the thermocline located at the site today, annually and during summer? If it's a winter signal as you argue, what is the difference between your values and the present winter temperatures (looks to be around 6 C according to WOA18)? Or even better, can you compare your temperature datasets to present day measurements from the same proxy in the same site/closeby? Same for alkenones (L543) and LDI?

A: The modern thermocline depth near the Hank site is ~30 m (Richardson and Pedersen, 1998; ICES J. Mar. Sci.). In winter the water column is more isothermal and in summer is more stratified (Richardson and Pedersen, 1998), so it is difficult to pinpoint what the integrated signal throughout the year would be. Nevertheless, water column and sediment trap studies indicate that the TEX86 signal in the sediments reflects that of the subsurface, i.e. 50–300 m, consistent with their role as ammonia oxidizing archaea (Church et al., 2010, Environ. Microbiol.; Schouten et al., 2012, GCA). GDGT-producing Thaumarchaeota also occur in deeper water layers (>300 m), however, they occur in lower abundances, and a clear mechanism that can explain their transport to the sediment is lacking (Wuchter et al., 2005; Paleoceanography). Given the shallow water column of the Pliocene North Sea, the influence of deep(er) dwelling organisms can be ignored. Regardless, a potential contribution of deeper dwelling organisms can be recognized by a high (>5) ratio of GDGT-2/GDGT-3 (Tyler et al., 2013; Glob. Planet. Change), which we calculate for our data and discuss on L529-535 of the discussion manuscript.

Present day TEX86 SST reconstructions for the North Sea range between 4.1 – 9.1 °C using the TEX86H calibration of Kim et al. (2010; GCA). These reconstructed SSTs correspond to observed SSTs of 10.3 – 11.3 °C (WOA18). Therefore, core top sediments in the North Sea are likely to underestimate the true SST; however, this statement may only be partly accurate insofar as the error on the proxy is ~2.5 °C (Kim et al., 2010). The Pliocene data from Hank hovers around ~9–12 °C for most of the record, which is in agreement or slightly higher than present day reconstructed SSTs using the same calibration. If it is interpreted as a winter signal as we argue, then the Hank Pliocene SSTs are approximately 3–6 °C higher than modern (van Aken, 2008; J. Sea Res.), which is in agreement with the data present in the manuscript in section 4.2.

Present day UK37 SST reconstructions for the North Sea: No suitable sites were found in the compilation of Tierney and Tingley (2018) of Uk37 recorded in surface sediments. The North Sea is a relatively underrepresented area in the UK37 calibration. However, reconstructed SSTs using alkenones from surface sediments from the Skagerrak region near the opening to the Baltic Sea record SSTs of 10–12 °C, approximately 1–2 °C higher than observed annual SST, and resemble those of May-June SST more (Blanz et al., 2005; GCA). Thus there is evidence for Uk37 recording summer temperatures (coinciding with bloom periods of haptophytes) in the circum-area North Sea.

Unfortunately, the core top calibration for the LDI SST proxy does not include core tops from the North Sea region. Therefore, we cannot compare our record to present day estimates of North Sea SSTs using the LDI proxy. This is due to the fact that the LDI is a relatively new proxy (Rampen et al., 2012; GCA), so it is not yet widely applied, especially not in combination with TEX86 and the Uk37. Indeed, the dataset presented in this manuscript is among the first paleo multi-proxy application studies of this proxy.

We will include a version of this discussion and comparisons of proxy-derived SSTs to modern SSTs in the revised version of the manuscript.

L540: Can you show a comparison towards some of these other records? All of these would relate to bottom water temperatures at your site—again, even if it is shallow there will be a clear difference between the top surface and the bottom water, especially if there is any seasonal biases. What does the temperature profiles look like today?

A: The temperature estimates of Wood et al (1993), Kuhlmann et al (2006), Johnson et al (2009) and Williams et al (2009) are not records of temperature over the whole of the Pliocene or mPWP intervals, but rather represent 'snapshots' of temperatures within these intervals. Thus it is not possible to plot our SST records and their reconstructed temperatures to directly compare SST evolution. However, it is possible to annotate Figures 2 and/or 6 in order to show the reconstructed temperatures based on these references in order to compare absolute values of temperatures reconstructed using the various proxies for approximately the same time interval. We will annotate Figure 6 to include annotations to the reconstructed temperatures of Wood et al (1993), Kuhlmann et al (2006), Johnson et al (2009) and Williams et al (2009), or a select few of the temperature estimates.

L550: Okhotsk and Rhode Island are quite different settings from your site. Are there any more local relevant studies to take into account?

A: We assume that by 'relevant' the reviewer means closer to the study site. In the revised manuscript, we will discuss the study of Blanz et al. (2005; GCA), who determined Uk37 values in a transect from the North Sea to the Baltic Sea and determined that for the Baltic Sea proper, there was no relationship between the alkenone UK37 indices with SST, whereas only the samples from Skagerrak plotted within 1 C of the global calibration of Müller et al. (1998; GCA). We will therefore add this discussion to section 4.2 in addition to these two sites.

L624: What is the present ocean circulation regime of the area? Should be presented in the introduction part of the paper. How large a fraction of the NAC enters the North Sea through the English Channel relative to north of Great Britain? Given the Pliocene geography of the area how different do you expect the circulation regime to have been?

A: The ocean circulation regime of the North Sea Basin today is dominated by wind-driven processes. The prevailing westerly winds on the north-west European shelf lead to a cyclonic anti-clockwise circulation, with the main input of water into the North Sea being from the north (Sündermann 2003; Oceanologia). We will add this information into the introduction of the paper. In the modern system, the modelled transport estimates from the HYCOM model indicate that the mean English Channel inflow is only 0.16 Sv (1 Sv = 10⁶ m³/s) versus the total mean inflow at the northern boundary of 2.22 Sv (Winther and Johannessen, 2006; J. Geophys. Res.: Oceans). In addition, the English Channel inflow varies intraannually, being weaker in winter and stronger in summer (Winther and Johannessen, 2006), meaning that the English Channel does not represent an important input of Atlantic water into the North Sea in the modern day. Given that the English Channel may or may not have been totally established in the Pliocene (Funnel, 1996; Quat. Sci. Rev.), and more than certainly closed during the M2 event, the NAC inflow would have originated entirely from the northern boundary during the Pliocene. See also discussion in the author response to the comment posted by Stijn de Schepper. Scant evidence for a Pliocene connection of the North Sea to the North Atlantic via the English Channel is observed in the Pliocene-age Coralline Crag fauna and flora, which indicate planktonic elements, as well as bryozoan-dominated deposits that bear resemblance to modern deposits (Funnel, 1996). We will include a brief discussion on this observed evidence for and against a Pliocene connection from the North Sea to North Atlantic via English Channel in the revised manuscript.

L639: Why? How? The full section here where you link the Zagwijn data to your data without showing them and without really making it clear if you can or cannot do this seems speculative to me.

A: We tentatively correlate the Zagwijn Taxodium-type curve to our MAT record in the text, however since the Pliocene stages defined by Zagwijn are probably incomplete (as we state), we cannot be 100% definite in our correlation. A more solid conclusion that we do make in the revised manuscript is the match between the stratigraphy concept of Zagwijn and the Hank data. We will include a brief description and discussion on this in the revised manuscript.

L652/L656: Can the variability be linked to the colder and warmer isotope stages? Does your variability compare to the changes seen in these other records?

A: We will explore the possibility that the variability in the SST biomarker records may be related to colder and warmer isotope stages in the global stack and in the local oxygen isotope record at Hank.

L661: The variability discussed by Bachem et al., 2017 and linked to gateway changes in not related to mPWP.

A: Thank you for clarification. We will remove the reference to Bachem et al. (2017) in this instance.

L663: The freshwater influence suggested for Clotten et al. 2018 cannot be used as direct support for fresh water influence at your site.

A: We understand the reviewer's point, however L661-663 outline a possible reason that Clotten et al. (2018) determined as a contributor for the high variability in their Uk37 SST record in Iceland. We did not mean to translate their explanation as evidence for sea ice-derived fresh water influence at our site in the North Sea. To clarify, we will modify the sentence to read: 'The high variability of Uk³⁷ SSTs at the Hank Site during the Pliocene is most likely due to a combination of factors, including the shallow depth of the SNSB, potential changes in the direction and strength of the NAC, and varying freshwater influence.'

L684-687: What is the argumentation and reasoning behind this statement? It is not clear how you support this conclusion.

A: See comments and author response to Stijn de Schepper, and our response to earlier comments on the possible influence of the NAC. We have decided to re-evaluate the statements we make regarding the influence of the NAC in the North Sea Basin as a result of these comments. We will revise this sentence in the revised manuscript and make the potential influence of the NAC on our site more tentative. It will read: 'Regardless, the high variability and warming trend in two out of the three organic SST proxies in the Pliocene North Sea indicate that the area was very sensitive to environmental changes, of which the specific climatic drivers remain unclear.'

L703: If this is the case you should show it. All the records you refer to are available online and can be plotted together with your data to document this statement. I would also like to see this relative to LR04 and your own d18O record, with the individual isotope stages visualized.

A: We will endeavour to obtain the SST data in order to plot a comparison of the variability of the records of Naafs et al. (2010), Bachem et al. (2017), and Lawrence et al., (2009), with our SST Uk37 record from the Hank site in order to show the similar variability in SSTs over the Pliocene. We will

follow the suggestion to make a figure with the SST Uk37 record from Hank relative to the LR04 and the d18O record from Hank with the individual isotope stages visualized.