Comments from the authors:

- We have found the comments from both reviewers very useful. We addressed them below.
- In addition to this, we have updated some figure captions. We have also updated and increased the resolution of some of the figures in the Supplementary file.
- We can move the raw data from the excel file into the PANGEA database when the manuscript is accepted
- The DOI for the model data will be activated once the manuscript is accepted.
- The reference list of the Supplement file is adjusted to the journal standard, as requested.

Replies to the reviewers:

Report 1 (Baatsen)

Referee comment on 'Sea surface temperature evolution of the North Atlantic Ocean across the Eocene-Oligocene Transition' by Śliwińska et al. In light of previous comments, the authors have done a good job addressing most issues while sticking the important information. It is now clear that this paper focusses on the proxy record of North Atlantic SST, trying to explain some of their observations to the results of climate models but leaving a more in-depth discussion of the latter outside of its scope.

Language has been improved, with some errors remaining mostly due to textual changes made that should be resolved in a final sweep. The main issue remains with readability/clarity of some of the figures, which I feel are not addressed properly. Some of the choices were explained in the authors' response, but little to no adjustments were made;

- Figure 4: there is too much information stacked on top of each other, using different colours, boxes and lines. I don't see much added value showing present-day SSTs (unless this serves as e.g. a model reference), yet using this over an anomaly with respect to PD uses about half of the figure. As it is now, it is also not clear to me what the main message is to the reader; the difference between the time intervals, with respect to the present, or rather which model simulation performs best? **We have improved the figure as follows:**
- the color palette match with the model output in figure 5 - it is a very good point regarding better colour match between figures!
- the main message is to compare the various model outputs with the proxy data – for this purpose it makes more sense to keep both: the model outputs and the proxy-data records.
- the modern SST are shown as a reference, so we would like to keep it. However, we have toned it down, by reducing it to a single line, so it is not dominating the figure.

- Figure 5: is a nice addition, but many of the trends are obscured by the large shifts during the asynchronous coupling phase. Consider leaving out the first ~3300 years, or splitting the figures into 2 parts adjusting the vertical axes. For consistency, it would be nice to colour-match the simulations throughout the different figures. **We have left out the first 3000 years and adjusted the size of each of the time-series. The color code for each model output is applied also in Fig. 4 for better synergy between figures.**

- Figure 6 (was 5): This is indeed a useful figure, but it does not serve its intended purpose well because of the very subtle colour scale and large range. In its current format, a certain proxy SST can easily be off by 4C or more and barely be visible. I suggest using a different colourmap and/or narrower temperature range such that the differences between both proxy and model SSTs are clear for the different scenarios. **This is a valid point. We have applied a colour-blind friendly palette, which is rather limiting. Other colour palettes did not improve the figure significantly and were not colour-blind friendly. Therefore, instead we added the SST data points for each site (pre- and post-34.5 Ma) on top of the temperature-colour scale. We believe that it makes the figure easier to read.**

- Figure 8 (was 7): I agree that the BSF is a very useful measure and that it should be depth averaged. Maybe my point was a bit unclear and was rather meant to take care in interpreting what is shown. The discussion of this figure mainly aims to explain part of the SST changes through current changes and the extent of SP/ST gyres. As the gyre contributions are likely quite weak at this point in the North Atlantic, the differences between both figures are probably mostly AMOC transports. Therefore, the link between SSTs and changes in depth-integrated transports as well as gyre extent is therefore in my mind not easily made based on the results shown.
We agree that the SST changes are not strictly wind-driven gyre changes because of the contribution to the AMOC. However, the SST increase when closing the Arctic has a very sharp front that sits right on the core site, and it corresponds also the boundary of the barotropic streamfunction. It makes therefore sense that the SST changes are related to the circulation changes (as there is no other obvious explanation for such a sharp SST front in the SST anomaly (see Fig7a, b, e). We now make it clear that this is not strictly related to the wind-driven gyre. In the discussion of Figure 8, we now note explicitly that: " The barotropic streamfunction in the model, a combination of the wind-driven gyre transport and the meridional overturning streamfunction, suggests ... ".
We further say: " with the streamfunction-derived Subtropical gyre" instead of just Subtropical gyre" to clarify it is not strictly the same.
And finally, we write that: "The position and strength of the gyres are likely model dependent and is here affected by the AMOC,"
These changes should clarify our point that the ocean circulation in the region (from winds and AMOC) is highly dynamic so could easily influence the SST there.

Minor comments:
- L280 and following: SST proxies are still referred to as observations here. **It is corrected now**

- L360: This is confusing; insufficient polar amplification would mean too cool rather than too warm polar temperatures? Figure 5 also shows that the simulations are underestimating rather than overestimating high latitude temperatures, in contrast to what is mentioned here. Additionaly, ‘polar’ temperature is somewhat of an unlucky term, as all of the proxies considered are near or equatorward of the Arctic Circle. Use ‘middle/high latitude’ instead?
Thanks for catching this error. This should be "too cold polar temperatures" of course. Also, we have changed throughout the manuscripts the description of the "polar" temperatures to high latitude temperature as we agree this is more descriptive of where these proxy locations are.

- Section 5.2: The authors suggested to add further information on the meridional overturning stream functions and SST of deep water formation regions. Much of this section was adjusted and a figure showing AMOC timeseries has been added. If there is any additional information and/or analysis in the supplement, there is no mention or reference here.
We added the meridional overturning stream functions series and deep ocean temperatures that we suggested to add in Figure 5. This concludes the additional analysis - as we explained, we want to avoid turning this too much into a modelling paper so only addressed the aspect directly related to the data.

Reviewer 2:

The authors have made substantial changes to address the concerns raised by reviewers. These changes include adding more detail about the uncertainty/limitations of both proxy data and model simulation, as well as adding and revising figures to improve clarity. I also appreciate that some paleoclimatic interpretation has been toned down or caveated accordingly. The authors have added some new text outlining the limitations of their model runs which were not in equilibrium. This seems like a critical issue but I do not have the expertise to assess. Overall I do find the manuscript much improved. However, there are still some technical issues and unclear reasoning that should be resolved. I also think that some minor restructuring of the manuscript might help further improve its

clarity.

[Line numbers as in the tracked changes version of the manuscript]

Line 40–44: These two sentences give off the impression that it is not possible to draw any firm conclusions from the results because the data reported are uncertain and model runs were not in equilibrium. This might lead to some readers thinking what is the point of this paper. I think the results presented are more valuable that that implied by these two sentences. Perhaps rephrase it to something more positive along the line of "future work"?
We agree and have now rephrased the abstract ending to conclude more positively as follows: " We conclude by highlighting remaining uncertainty in various aspects of proxy and modelling work which could be improved to shed further light on the processes responsible for the cooling trends identified here prior to and across the EOT in the high latitude North Atlantic."

Related to this, a dedicated section on future work / suggestions on what else to do to improve our understanding of temperature evolution across EOT might improve the structure and clarity of discussion. I would also shorten the conclusions by moving all the suggestions / future work to this proposed new discussion section.
The uncertainties related to the conclusions are a result of data scarcity and uncertainty, as well as model uncertainty and the fact that these simulations explore only very few of the possible EOT scenarios. Any model or data study is likely to help with, but none to solve alone the question of the EOT. We have thus concluded with a general recommendation that more proxy SST data and modelling results are needed. However, we have now added at the end the suggestion of a formal EOT model intercomparison project.

Line 164: "where 37:3 stands for methyl ketone and 37:2 for alkenone". Both are alkenones, which are ketones. The numbers refer to the number of carbon atom and double bonds in the molecule, respectively. Please revise accordingly. Also the notation [C37:2] is more common that 37:2.
We have corrected the text as suggested.

Line 170: "…has been present only for 270 kyr." 270 kyr during which time interval? Suggested change "the past 270 kyr". Also a reference is needed here.
We have corrected the text as suggested.

Line 175: "…overlapping with the calibration used for E. huxleyi." Specify which calibration; several calibrations derived from E. huxleyi have been reported, some of which are very different from the Müller calibration. Would be also helpful to briefly explain why similarity with a culture calibration confirms the robustness of the core-top calibration.
We have added a relevant reference. See also our response to the comment below. We think that any further explanations discussing regarding similarity with a culture calibration are out of the scope of this paper.

Line 175–177: The Müller calibration is based on data from the global ocean, so it is not true that it "mostly uses surface sediment data from the North Atlantic Ocean…". That would be the Rosell-Mele et al 1995 GCA Atlantic calibration.
This is a valid point, which we considered. We have rephrased the sentence as follows the calibration of Müller et al., (1998) is near identical to the culture-based calibration used for *E. huxleyi* by Prahl et al. (1988) and is the most commonly used calibration for the UK37'-derived SST calculation of the late Paleogene to Neogene strata in the northern high to mid latitudes (see e.g. Liu et al., 2009; Herbert et al., 2020; Weller and Stein, 2008)."

Line 225: Fig S2 shows temperature records based on different calibrations, not calibrations (ie regressions) themselves. Please rephrase. **Good point. It is corrected now.**

Line 231–232: Unclear reasoning. Just because proxies are based on organisms with different

ecological preferences does not mean that they are associated with uncertainty. Do the authors mean inter-proxy comparison is associated with uncertainty? **We mean that each of the proxy not perfectly reflecting annual mean SST (i.e. 0-30 mixed layer) but is generally reflecting different seasons and different depths, and therefore carry different uncertainties with respect to reconstruction annual mean SST. The sentence is rephrased to: "These two proxies are based on organisms with different ecological preferences and thus may reconstruct temperatures of different seasons and in particular depth compared to each other." And in the given context we believe that the meaning is now clearer.**

Line 232–234: As I commented in the previous round of review, the reported UK'37 and TEX86 records (the entire time series) are really not that similar in trends nor absolute values; the UK'37 record shows a strong cooling trend while TEX86 record shows very little change. The match is only true for the early Oligocene. I would therefore be specific with the time interval during which proxies match. **We have stated that the similarity in trends is in the given interval "(covered by the interval from ~240 to ~190 mbsf), line 221). The magnitude of cooling at ~183 m depth is different between the proxies, but it is observed in both proxy records. Also, as shown on Fig. S2B the trends are fairly similar. As stated above, this is to be expected from proxies which are from organisms with different ecological preferences.**

Another piece of information regarding surface vs subsurface origin of TEX86 signal is the GDGT-2/GDGT-3 ratio (Taylor et al 2013 GPC), which is used by some workers to indicate deep-water archaeal contribution to sedimentary GDGTs.
The GDGT-2/GDGT-3 ratio was calculated for all samples and is shown in our supplementary file. The ratio is low (< 5) and rather stable, suggesting no major subsurface contribution to sedimentary GDGTs.

Line 280: I would remove "Observations of". Strictly speaking proxy-derived SST estimates are not really "temperature observations". **Good point, we have removed word "observation" from lines 285 and 286, and changed to: "4. Proxy-derived sea surface temperature" and "4.2. Sea surface temperature in the North Atlantic across the EOT".**

Line 287–288: Unclear what is meant by this sentence. Some elaboration would be helpful. **We have rephrased the sentence to make our intentions clearer.**

Line 320: "Proxy-derived" instead of "Observed". **It is corrected**

Line 408: "latter" not "letter". **It is corrected**