Authors Response to review comments for:

Deep Ocean Temperatures through Time

We thanks the reviewers for their comments, and have responded to all points. However, a few points are difficult to respond to within the context of these simulations. The simulations represent a huge amount of CPU time (approximately 1 million years of model simulation) requiring a significant fraction of a large HPC for about 9 months. It is therefore not practical to perform several further sets of simulations to investigate aspects, such as whether there are multiple equilibria. We therefore have had to amend our text to highlight the limitations of the work.

Similarly, we propose to produce a supplementary set of figures which are an atlas of the simulations. One page per time period (109 in total), showing the geographical maps for paleogeog, surface temperature, deep ocean temperature, mixed layer depth. These will address some of the requests from the reviewers.

Response to Reviewer A

(Grey text = reviewers comments, Black text= our response, Red text= addition/changes to paper)

Major comment:

"However this paper could become far better with an extended discussion. Here authors conclude that the change in the deep water temperature should not be taken as representative of the globally change of surface temperature for Mesozoic-Paleozoic eras.
However the discrepancy between data and models is mainly driven by two warm periods (Jurassic and Triassic) where deep waters were formed in shallow sea at low latitudes (lines 517-519). This result is very interesting and challenges the initial suggestion of Emiliani (1954) that deep water continues to be formed at high latitudes in climate much warmer than today's. By adding a few diagnostics, the present study may become a key paper to understand the origin of warm deep waters and which conditions are required to form warm deep water at low latitudes (as initially proposed by Chamberlin, 1906). By identifying conditions for formation of deep waters in very warm climates (Cretaceous, Triassic), the scatter between deep ocean/mean surface temperatures should be significantly reduced for pre-Cenozoic periods. Since this result may substantially change the conclusion of the present paper, I recommend a major revision."

We are not the first to suggest that there was deep water from the tropics. As discussed in lines 391 onwards, Brass et al 1982 was the first to suggest that the hypothesis of deep water always forming at high latitudes was not necessarily correct for all paleo times. Using a simple convective plume model, Brass et al showed "that the depth of termination and horizontal spreading of the plumes is directly related to their buoyancy flux which is the product of the density difference between the plume fluid and the surrounding environment and the volume flux of the plume. The plume with the greatest buoyancy flux will form the bottom water and plumes of lesser buoyancy flux will terminate and spread out at intermediate depths". Hence the key issue is where is the strongest

buoyancy flux. Brass et al then suggested that this could occur in warm conditions in the tropics, particularly if there was significant epicontinental seaways.

Similarly, we are not the first climate modelling group to show that this does occur.

The reviewer states that "…the discrepancy between data and models …". There is no discrepancy because, as far as I am aware, there is no data to show where deep-water is formed in the Jurassic and Triassic. Any discrepancy is between conceptual hypothesis (of Emiliani etc.) and the computer models presented here which follow the detailed laws of physics, such as Newton's laws and the laws of thermodynamics.

Thus none of our results and figures change depending on the interpretation (we cannot change the laws of physics). However, the reviewer does highlight our need to better explain our results, so we have expanded the paragraph discussing warm deep water, and included two extra figures. Specifically we have replaced the final text from line 391-394 by:

The idea that deep water may form in the tropics is in disagreement with early hypothesis (e.g. Emiliani (1954)) but has been previously suggested as a mechanism for warm Cretaceous deep water formation (Brass et al., 1982). Deep water typically forms in convective plumes. They showed that the depth and spreading of these plumes is related to the buoyancy flux with the greatest flux leading to bottom water and plumes of lesser flux leading to intermediate water. Brass et al (1982) suggested that this could occur in warm conditions in the tropics, particularly if there was significant epicontinental seaways and hypothesised that it "has been a dominant mechanism of deep water formation in historical times". It is caused by a strong buoyancy flux linked to strong evaporation at high temperatures.

Our computer model simulations are partly consistent with this hypothesis. The key aspect for the model is a relatively enclosed seaway in the tropics and warm conditions. The paleogeographic reconstructions (see supplementary figure 1) suggest an enclosed Tethyan-like seaway starting in the Carboniferous and extending through to the Jurassic and early Cretaceous. However, the colder conditions of the Carboniferous prevents strong tropical buoyancy fluxes. However, when we get into the Triassic and Jurassic, the warmer conditions lead to strong evaporation at low latitudes and bottom water formation in the tropics. This also explains why we see more tropical deep water (and hence poorer correlations between deep and polar surface temperatures in figure 6) when using the Foster et al CO_2 since this is generally higher (and hence warmer) than the smoothed record.

An example of the formation of tropical deep water is shown in figure xx. This shows a vertical cross-section of temperature and salinity near the equator for the Ladinian stage, mid-Triassic (240Ma). The salinity and temperature cross-section clearly shows high salinity warm waters sinking to the bottom of the ocean and spreading out. This is further confirmed by the water age tracer, figure yy. This shows the water age (measured as time since it experienced surface conditions, see England 1995) at 2731m in the model for the Permian, Triassic, Cretaceous and present day. The present day simulation shows that the youngest water is in the N. Atlantic and off the coast of Antarctica, indicating that this is where the deep water is forming. By contrast, the Triassic period shows that the youngest water is in the tropical Tethyan region and that it spreads out from there to fill the rest of the ocean basin. There is no young water at high latitudes, confirming that the source of bottom water is tropical only. For the Permian, although there continues to be a Tethyan-like tropical seaway, the colder conditions mean that deep water is again forming at high latitudes only. The Cretaceous is more complicated. It shows younger water in the high latitudes, but also shows some young water in the Tethys which merges with the high latitude waters. Additional indicator of the transitional nature of the Cretaceous are the mixed layer depth (see supplementary figure xx). This

is a measure of where water is mixing to deeper levels. For this time period, there are regions of deep mixed layer in both the tropics and high latitudes, whereas it is only deep in the tropics for the Triassic and at high latitudes for present day.

This mechanism for warm deep water formation has also been seen in other climate models (e.g. [Barron and Peterson, 1990]). However, Poulsen et al 2001 conclude that in his model of the Cretaceous high-latitudes sources of deep water diminish with elevated CO_2 concentrations but did not see the dominance of tropical sources. Other models (e.g. Ladant et al 2020) do not show any significant tropical deep water formation, suggesting that this feature is potentially, a model-dependent result.
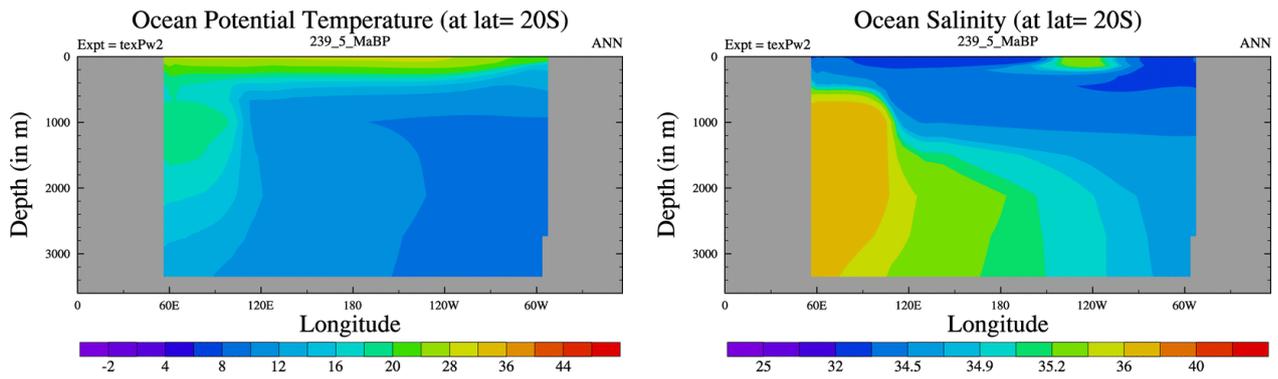
Figure xx. Longitudinal cross section at 20S of ocean potential temperature and salinity for the Ladanian (240Ma). Temperature is in C and salinity is in PSU.
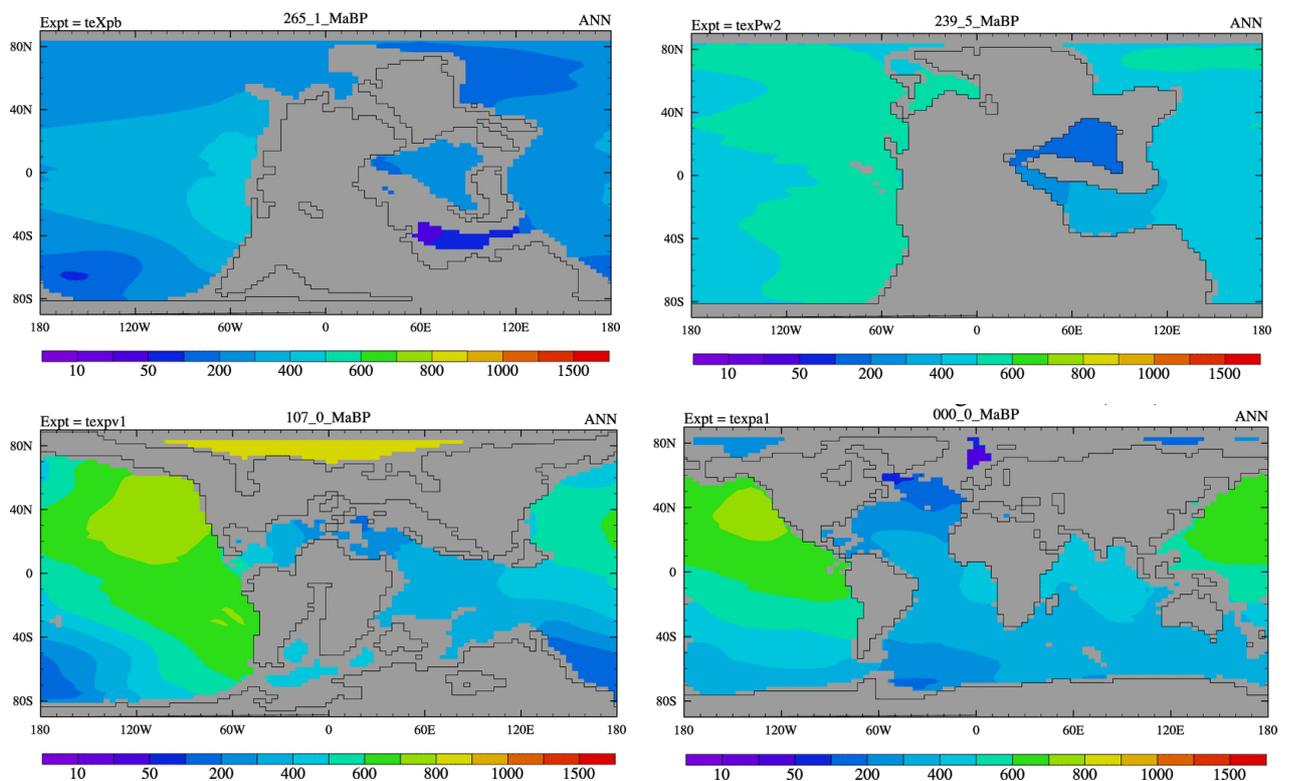
Figure yy. Modelled age of water tracer at 2731m for 4 different time periods (a) 265Ma, (b) 240Ma, (c) 107Ma, and (d) 0Ma. Units are years.

Minor Comments:

lines 381-395 (section Correlation of Deep Temperatures to Polar Sea Surface Temperatures) Here, the authors assume that most of the deep water in the model is formed
at the surface at high latitudes (which explain why they explore the polar amplification
later). However HadCM3 seems to be able to form dense warm saline bottom water insubtropics when the Earth's climate becomes very warm. This process likely explains why purple dotes (fig. 6 and fig.9) are clearly above the proposed slope and why polar surface/deep ocean temperatures (fig.6) and deep ocean temperatures/global temperature (fig.9) are not well-correlated.

This is a very similar comment to the major comment and we believe we have addressed the scientific issues in the previous response. The model does not assume anything about the location of deep water formation. This is a product of the laws of physics, as implemented within our model. We are plotting figure 6 to show that the "normal" assumption of the link between deep water and surface polar waters is not fully consistent by our physics-based model. We agree with the reviewer those time periods which show the largest departures are because of the deep water site, and we believe we have addressed the reasons for this in the revised text above.

We also note that the reviewer asks us to "reconstruct the Earth temperature (for the whole Phanerozoic) using deep water temperatures assuming that "deep ocean water does not always form at polar latitudes". This is impossible to directly address with a computer climate model. We cannot turn off/on where deep water forms. However, we do correlate the deep ocean water temperatures to global mean surface temperatures (see section starting at line 467 and figures 9 and 10). This confirms that there is a large time interval during the Paleozoic to the Jurassic when the errors at reconstructing global surface temperature from deep ocean temperatures are large.

The reviewer is correct that Poulsen did not show global changes driven by the tropics, but other models have. We have changed the text (see above).

models and data. If we assume the data has a seasonal bias, and select the summer seasons : : :. reduced by 4_C" This sentence implies that deep waters in polar regions are formed during the summer season (which means that the cause for buoyancy loss is salinity (without the formation of sea-ice)! Moreover this sentence is inconsistent with correlations made later in the paper. For example the caption of the figure 6 considers the polar temperature averaged in winter. This sentence should be removed and the discrepancy between models and data more discussed in the manuscript (see comment above).

In this section, we are comparing our surface temperatures to planktic foram reconstructions of SST, and are not discussing deep water data or mechanisms. The planktic data is potentially likely to be summer biased. We are not making any statement about deep water formation at this point. We have added a sentence to clarify this.

Of course, in practice deep water is formed during winter but the observed planktic data does not precisely record winter temperatures.

- lines 508-513 (section Discussion and Conclusion) Paleogeographies are often mentioned as the main reason of the results outlined in the study. Unfortunately the direct effect of this factor is not well illustrated. Since the polar amplification depends on geographies used in this study, a figure showing the Poleward Energy Transport (or/and a paragraph) could be included in the section devoted to the " polar amplification"

This is a really good point but highly non-trivial. The patterns of change in heat transport are complicated, and depend on numerous aspects of the paleogeography and other aspects of the model and we cannot identify any simple relationships. Whole papers have been written on changes in energy transport for one time period, let alone for 2 x 109 simulations presented here. A detaile analysis will be performed in a separate paper. For the moment, we have just highlighted the complexities in the following new paragraph, in the polar amplification section.

Changes in heat transport also play a potentially important role in polar amplification. Examination of the modelled poleward heat transport by the atmosphere and ocean shows a very complicated pattern, with all time periods showing the presence of some Bjerknes compensation (Bjerknes, 1964, and see, for example, Outten et al, 2018 for example in CMIP5 models). Bjerknes compensation is where the change in ocean transport is largely balanced by an equal but opposite change in atmospheric transport. For instance, compared to present day, the mid-Cretaceous and Early Eocene warm simulations shows a large increase in northward atmospheric heat transport, linked with enhanced latent heat transport associated with the warmer, moister atmosphere. However, this is partly cancelled by an equal but opposite change in the ocean transport. E.g. compared to present day, the early Eocene northern hemisphere atmospheric heat transport increases by up to 0.5PW, but the ocean   transport is reduced by an equal amount. The net transport from equator to the N.Pole changes by less than 0.1PW (i.e. less than 2% of total). Further back in time, the compensation is still apparent but the changes are more complicated, especially when the continents are largely in the Southern hemisphere. Understanding the causes of these transport changes will be the subject of another paper.

- section 2.4 The authors implicitly assume that the initial state for the ocean has a marginal effect (which means that the final state is always the same whatever initial conditions used, so there is no hysteresis). However this point may have importance. Indeed, because late Jurassic deep waters are warmer than for the Cretaceous, a sensitivity test should be performed using temperature profiles obtained for Jurassic

We explicitly acknowledge that our results could be dependent on the initial conditions (line 226). The existing simulations already account for a huge amount of computer time and it would be impractical to investigate the sensitivity to initial conditions for all time periods. This is because there is no systematic way to locate hysteresis without performing multiple simulations for each time period. Simply initialising a Cretaceous model with Jurassic temperatures is not rigorous. We might show a transition, but we might not but the latter would not show that we do not have hysteresis, it would simply show that this particular sensitivity test did not show it. We have instead extended the discussion to expand the following caveat of our work:

Although it is always possible that a different initialization procedure may produce different final states, it is impossible to explore the possibility of hysteresis without performing many simulations for each period, which is currently beyond our computing resources. Previous studies using HadCM3L (not published) with alternative ocean initial states (isothermal at 0C, 8C, and 16C) have not revealed multiple equilibria but this might have been because we did not locate the appropriate part of parameter space that exhibits hysteresis. This remains a major caveat of our current work and which we wish to investigate, when we have sufficient computing resource.

This is shown in figure 5, where the final two images (d and e) are the modern model and Levitus observed data. However we did not describe this result. We have added a sentence to correct this.

Figure 5 also shows the modelled deep ocean temperatures for present day (Fig5d) compared to the World Ocean Atlas Data (fig5e). It can be seen that the broad patterns are well reproduced in the model, with good predictions of the mean temperature of the Pacific. The model is somewhat too warm in the Atlantic itself, and has a stronger plume from the Mediterranean than is shown in the observations.

The figure caption was incorrect.

Our mistake. Now corrected.

Response to Reviewer 2

We have added additional lines to show the benthic temperatures in Pacific and Atlantic basins separately.

We reiterate that the purpose of the comparison with data is to demonstrate that the model is broadly consistent with data but not a detailed validation. This is because the comparison is strongly influenced by the selection of CO2 and there is considerable uncertainty in this. We are currently undertaking a large 2-year project to perform a much more detailed comparison to data throughout the Phanerozoic, including the role of CO2.

2) Because the analyses largely focus on benthic temperature, I think benthic temperature is a better metric for determining equilibrium than global ocean temperature.

The metric we use is the volume integrated global ocean temperature and this is dominated by the slower time scales of the deep ocean and hence is a good metric, even for deep water. We have also examined temperatures averaged over the bottom 10 levels of the model, and temperatures at 2116m. They all show the same shape. We have added text at line 240.

The volume integrated temperature is dominated by the deep ocean so trends are similar if we look at the deep ocean only (e.g. at 2116m) or the average over the bottom 10 layers of the model.

Related, there is unexpected variability in some of the long-term temperature responses in figure 3. Can you rule out this type of response in the simulations that you only ran for 2000 years?

Those models that showed long-term behaviour (beyond 2000 years) failed our multiple criteria at year 2000, specifically although the trends might have been small, the energy balance was not. Hence we are reasonably confident that those models that passed all criteria at year 2000 were near equilibrium.

The strength of using multiple constraints is that a simulation may, by chance, pass one or two of these criteria but were unlikely to pass all three tests. For example, all of the models that we extended failed at least two of the criteria.

3) Additional details above how the "smoothed" CO2 curve was derived are necessary. It sounds like it is based on temperature reconstructions, which makes its better agreement with the benthic records unsurprising. The information on lines 174-177 is not sufficient.

We have added (from line 176):

Specifically, over the last decade and using proprietary paleogeographies, we performed multiple simulations at different CO_2 values for several stages across the last 440 million years and tested the resulting climate against proxy data (Harris et al., 2017). We then selected the CO_2 that best matched the data. For the current simulations, we linearly interpolated these CO_2 values for every stage. The resulting CO_2 curve happens to look like a heavily smoothed version of the Foster curve, but it was derived from a very different approach.

4) Why use a linear relationship for reconstructing surface temperature from deep ocean temperature? The relationship in figure 9 is non-linear. Might a non-linear fit improve the model based reconstructions?

It might improve the overall fit but there is relatively little justification for doing so. Except for the very coldest temperatures, the residuals of the linear fit are near Gaussian suggesting that there is no systematic evidence for non-linear variations. When studies use deep ocean temperatures to estimate global surface temperatures, they always assume linearity. We have added a line:

When we examine the residuals from the linear fit, they are near Gaussian suggesting that a linear fit is appropriate.

5) How well does HadCM3 simulate present-day benthic temperature? If there are significant biases, maybe anomaly maps would make more sense. Also, a sentence about the dynamic vegetation would be nice.

See comments in response to reviewer 1. We also point the reviewer to Valdes et 2017 which shows the vegetation, and have added at line 85:

The paper also documents the performance of the dynamic vegetation model but not the deep ocean temperatures. These are discussed in section 3.1

Line 54 – Consider efforts using different models? (e.g. Donnadieu et al. (2016); Ladant et al. (2020))

Added.

Line 155 – Can we assume ocean salinity was constant through time?

NO but we have little knowledge of how it might have changed. Added sentence.
We have little knowledge of whether ocean salinity has changed through time, but a global change in salinity has relatively little impact on the ocean circulation.

Line 222 – What were the initial ocean conditions for the original simulation(s)?

This is explained at line 232

Line 226 – All simulations were run in sequence? If not, please provide additional details about your spin up procedure.

NO. If we ran the models in sequence, the work would have taken more than 30 years to complete! We have added text at line 234:

Simulation were run in parallel so were not initialised from the previous stage results. In total, we performed almost 1 million years of model simulation and if we ran simulations in sequence, it would have taken 30 years to complete the simulations. By running these in parallel, initialised from previous modelling studies, we reduced the total run time to 3 months, albeit using a substantial amount of our high performance computer resources.

Line 229 – Citation for the lack of multiple equilibria?

None. If we had found multiple equilibria, we would have published! There are no papers because we have never found them.

Agreed. The basis was that that trends and imbalances were smaller than typical orbital variability. Text added:

The numbers chosen were arbitrary but were to ensure that any trends and imbalances were smaller than typical trends associated with orbital variability.

Figure caption corrected (see response to reviewer A). The choice of 1000m was also a bit arbitrary but was to avoid including shelf data which is generally not included in the compilations of benthic data. Text added at 265:

to avoid continental shelf locations which are typically not included in benthic data compilations .

As mentioned in text, the data compilations can include data from ridges which are at 2000m in benthic compilations, and group all data together irrespective of depth (i.e. the data compilations essentially assume there are no vertical gradients between 2000m and the bottom of ocean). I selected the model level 2116m as a compromise. If I selected a lower level, e.g. at 4000m, then only a very small part of the ocean is sampled. If I used the bottom of the ocean, then it can be difficult to interpret since it includes a big range of depths. However, for all diagnostics, the difference between the bottom temperatures and the 2116m data were small. We have updated the text:

The observed data was collected from a range of depths - including mid-ocean ridges whose depth can vary from 2000m to the true bottom of the ocean. Thus benthic data compilations can be of variable depth and assume that the vertical temperature gradient is small. And "…, and hence throughout the rest of the paper we frequently use the model 2116m temperatures as a surrogate for the true benthic temperature. This is also useful because the area of deep ocean can change substantially for even deeper levels."

I have modified this line since it is difficult to find a reference (even though it is often discussed in workshops).

Both dips in the Foster-CO_2 simulations correspond to relatively low estimates of CO_2 concentrations. For these periods, the dominant source of CO_2 values for these periods is from paleosols (fig.1) and thus we are reliant on one proxy methodology. At other time periods, the paleosols are often at the lower end of estimates (Foster et al 2017)

I have added Goldner. I selected You et al because I think it was the first to show discrepancy. I was not attempting a comprehensive review.

Line 303 – Again, why use 2000 m temperature for comparison here?

See above.

Lines 309-313 – The proxy community would be very interested in a few more details on the evolution of deep water formation through time. Maybe a few plots of mixed layer depth would help show this?

Now included in supplementary information.

Line 322 – Again, comparison at proxy locations would be more informative.

See response to first comment.

Line 352 – Sluggish circulation can also modify salinity and ocean d18O (e.g. Zhou et al., 2008)
Modified text:
…, with weak horizontal temperature gradients at depth (though salinity gradients can still be important, Zhou et al 2008).

Line 355 – Upchurch et al. (2015) also tested cloud microphysics.

Reference added:
Line 372 – I'm surprised that the mixed layer depth compared worse than polar temperature with benthic temperature.

To be honest I was surprised too, but I think it is the complexity of the problem. The mixed layer depth is spatially, seasonally and duration varying in intensity and area. It was therefore difficult to precisely define the key sinking area, and to quantify the mass of water exchanged. A further study might attempt to do better quantification of the effects by outputting the mass exchange but we did not archive the variable in these simulations.

Line 377 – Please show the response with just Cenozoic data

The purpose of the paper is to model the whole of the Phanerozoic so we are reluctant to exclude other data. However, we have split figure 6 into all time periods (a), and Cenozoic only (b)

Line 395 – What happens to the high latitude temperature response when deep water forms primarily at the low latitudes?

There is no obvious change to the response, but to fully answer such questions would require some additional sensitivity tests.

Line 409 – Is HadCM3 underestimating the temperature gradient response or just underestimating sensitivity to CO2?

The results of EoMIP (Lunt et al 2012) and DeepMIP (Lunt et al 2020) suggest that it is the temperature gradient is the biggest problem. I have added these references to the text.

Line 434 – Can you speculate about what this means for proxy interpretation?

Not sure what the reviewer is thinking about here, but have added:

The difference between the southern and northern hemisphere response complicates the interpretation of the proxies and leads to potentially substantial uncertainties. This is discussed in section 3.4 and figure 10.

Line 440 – Isn't the albedo feedback somewhat offset by the fact that SSTs cannot get below 2_C?

The reviewer is partly correct but the ice albedo back does dominate.

Figures –

Need lettering on all panels. What each panel illustrates is not always clear.

Done

Figure 1 – Land fraction is very low before 250 Ma! How does this impact your results?

This is partly discussed in Farnsworth et al 2019 and in the surface polar amplification section (now section 3.4)

Figure 2 – I think this may work better as a supplement figure with more time periods.

We have kept it but added supplementary figures with all time periods.

Figure 3 – What do the colors represent? Maybe easier to see with thinner lines.

Used thinner lines. Expanded figure caption to explain that each line represents a different time period.

Figure 4 – Should be "less than 1000 m". What is the Bemis latitudinal correction? Why not use a salinity based correction? Any correction for Mg/Ca?

Corrected. The data is from Cramwinkel and they performed all of the corrections. I have clarified this in the figure caption. Reference added for Bemis (correcting for latitudinal gradient of del18O of seawater).

Figure 5 – Some bias in the present-day simulation at 2000 m, which should be mentioned in the text.

Done. See response to reviewer A

Figure 6 – Are SSTs calculated below the sea ice? Might help to plot regressions for Foster and Smoothed separately.

Yes. Both lines added.

There is negligible ocean crust remaining from this time period so reconstructing the depth of the ocean is impossible. It was arbitrarily set to 3680m depth. Added line to figure caption and to section 2.3 where we describe the paleogeogs.

Improved