

Authors Response to review comments for:

Deep Ocean Temperatures through Time

We thank the reviewers for their comments, and have responded to all points. However, a few of the more detailed points are difficult to respond to fully within the context of these simulations. Most published climate modelling papers tackle one time period, and maybe use one or a handful of model simulations. By contrast, we have 2 x 109 simulations using very different paleogeographies. Therefore it is impossible to go into the detail (e.g. why does the ocean heat transport change?) that a normal paper might target. To answer all such questions in depth would probably require 109 papers, one for each time period!

Similarly, 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 some of the suggested additional aspects, such as whether there are multiple equilibria. We therefore have had to amend our text to highlight the limitations of the work rather than carry out additional simulations.

We have produced 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 paleogeography, surface temperature, deep ocean temperature, and mixed layer depth. This will also address some of the requests from the reviewers.

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

Response to Editors Request

The editor request we address the issue of the improved closure of the water cycle in HadCM3, compared to previous simulations such as Farnsworth et al. (2019) and Lunt et al, 2016.

This is currently not possible to fully address because our modelling uses different paleogeographies and CO₂ values. Farnsworth is currently re-running his set of simulations and will be preparing a paper which will address this issue. We have added a couple of sentences:

We have not directly compared our simulations to the previous runs of the (Farnsworth et al., 2019a) because they use different CO₂ and different paleogeographies. However in practice, the increase of salinity in their simulations is well mixed and seems to have relatively little impact on the overall climate and ocean circulation.

Response to Reviewer A

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 formed in 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 time periods. 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 in their model.

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. The reviewer discusses Emiliani but they only discuss data from the Tertiary and we do not have tropical deep water during these periods. Any discrepancy is due to an (invalid in our opinion) extrapolation of Emiliani to all geological times.

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:

This mechanism has been previously suggested as a mechanism for warm Cretaceous deep water formation (Brass et al., 1982), (Kennett and Stott, 1991). The idea that deep water may form in the tropics is in disagreement with early hypothesis (e.g. (Emiliani, 1954)) but they were only considering the Cenozoic and our model does not simulate any low latitude deep water formation during this period. We only see significant tropical deep water formation for earlier periods, and this has previously been 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 figures) 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 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 fig. 8. 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, fig. 9. 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. An additional indicator of the transitional nature of the Cretaceous are the mixed layer depths (see supplementary figures). 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 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 CO2 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.

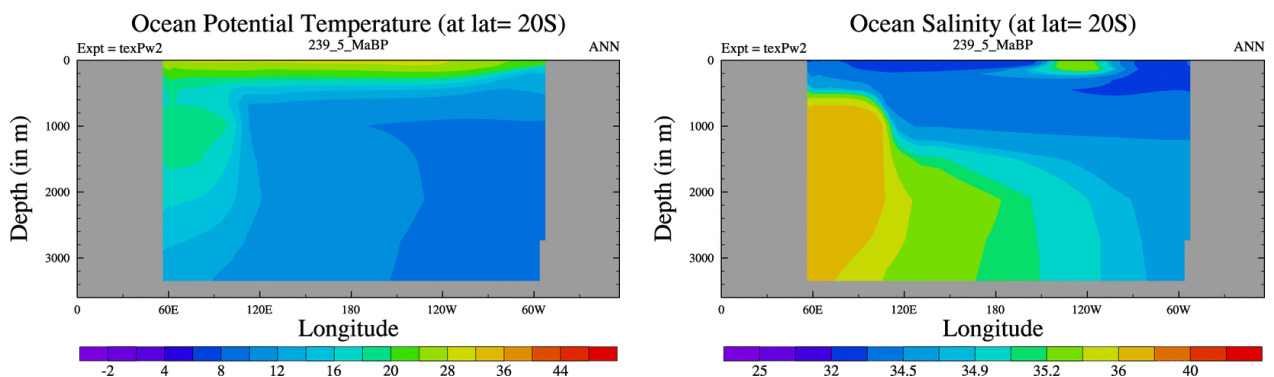
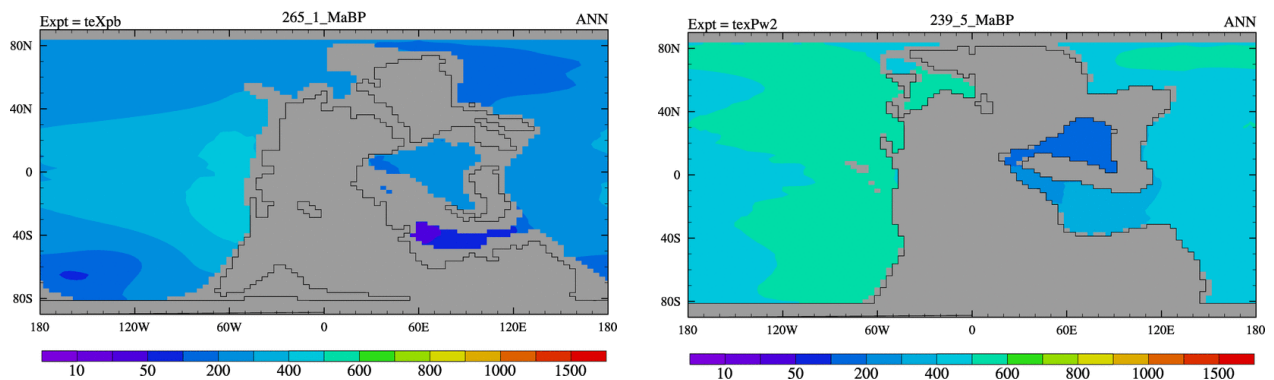


Figure xx. Longitudinal cross section at 20S of ocean potential temperature and salinity for the Ladanian (240Ma). Temperature is in C and salinity is in PSU.



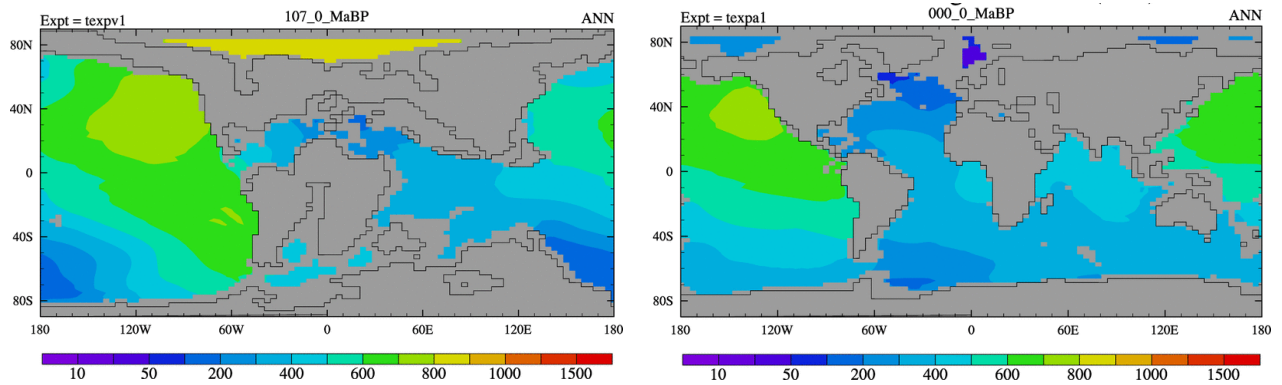


Figure yy. Modelled age of water tracer at 2731m for 4 different time periods (a) 265Ma, (b) 240Ma, (c) 107Ma, and (d) 0Ma. Units are years.

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 dots (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. Consequently this result seems to indicate that we need to distinguish two oceanic states for the formation of deep waters 1) when the main cause for buoyancy loss is salinity and 2) when the main cause for buoyancy loss is cooling.

This issue appears fundamental because this results implies that the change in polar surface temperature is not representative of the deep ocean temperatures. Consequently a "systematic" correlation between polar/deep ocean temperatures cannot be made. If the authors want to keep their initial conclusion, they have to demonstrate that the Emiliani's suggestion (1954) that most of deep waters continue to be formed at high latitudes is always true - even in a climate much warmer than today's. For solving this issue, the authors should: - reconstruct the Earth temperature (for the whole Phanerozoic) using deep water temperatures assuming that "deep ocean water does not always form at polar latitudes" (line 386-387). - explore the effects of shallow sea at low latitudes for warm periods (here Triassic and Cretaceous) to decipher why the formation of deep water is so different, in both cases.

I realize that the authors may not want to undertake the project I have outlined. In that case they should claim only a speculation, not a conclusion, and they have to rewrite the paper in taking into account this major issue.

References to add: On a Possible Reversal of Deep-Sea Circulation and Its Influence on Geologic Climates. Chamberlin. T. Proceedings of the American Philosophical Society Vol. 45, No. 182 (1906), pp. 33-43 Temperatures of Pacific Bottom Waters and Polar Superficial Waters during the Tertiary, C. Emiliani Science (1954), Vol. 119, Issue 3103, pp. 853-855 DOI: 10.1126/science.119.3103.853

This is a very similar comment to the major comment and we believe we have addressed the scientific issues in the previous response. In particular, we want to highlight in the strongest terms

that the climate model does not assume anything about the location of deep water formation. It is a product of the laws of physics, as implemented within our model. The key thing is that when previous authors have used benthic temperatures as a proxy for global mean temperatures, it is they who assume the linkage with polar surface temperatures (as we discuss in the introduction and also see Hansen and Sato 2012 and many related papers).

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 with our physics-based model. We agree with the reviewer that 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 or indeed from data. 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.

- lines 394-395: If Poulsen et al. (2001) has mentioned the formation of dense warm saline water in subtropics by the geographic isolation (Mid-Cretaceous boundary conditions), Poulsen’s simulations did not show a global circulation driven by deep water in the sub-tropics. Moreover, I do not understand why this result could be considered as model-dependant. Indeed both cases are generated by the same model.

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). We do not understand the final sentence - Poulsen uses a different climate model to our work.

- lines 349-355 (section 3.2: Comparison of Model Sea Surface Temperature to Proxy Data) I disagree with the following sentence - lines 351-352: “: : 15_C mismatch between 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 has potential 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 detailed analysis will be performed in a separate paper. For the moment, we have just highlighted the complexities in the following new paragraph, in the discussion section, and added a supplementary figure showing the complexity of transport.

Changes in heat transport also play a potentially important role in polar amplification. In the supplementary figure, we show the change in atmosphere and ocean poleward heat fluxes for each time period. 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) (see (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 (instead of values from previous model simulations) to initialize the Cretaceous ocean.

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. Furthermore, it is impossible to prove the absence of hysteresis, only that we have not found 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 or bistability 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 bistability. However, other studies have shown such behaviour (e.g. (Baatsen et al., 2018)). This remains an avenue for further research, which we wish to investigate when we have sufficient computing resource.

- section 3.1 To demonstrate that the GCM is well designed to compare deep ocean temperatures to benthic ocean data, the revised version of the manuscript should contain a validation test using modern conditions (deep water temperatures simulated versus modern data) - or at least a reference.

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.

- line 265 "less than 1000m" seems to be not consistent with the caption of the fig.4 (line 644)

The figure caption was incorrect.

- general shape of the manuscript The manuscript is organized by headings and subheadings from pages 2 to 10 but not after, why ?

Our mistake. Now corrected.

Response to Reviewer 2

1) It would be useful to compare the proxy data with model data at the proxy collection sites instead of global averages. This could provide additional information about how well the current benthic records can reconstruct deep ocean temperature through time and how well the benthic records reflect surface temperature.

We reiterate that the purpose of the comparison with data is to demonstrate that the model is broadly consistent with data but is 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 point-based 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.

Due to the colossal scale of our set of simulations (in total we have performed approximately 1 million years of simulation), we were unable to archive the full 3-D structure of the ocean and atmosphere for all years. We therefore have timeseries for selected variables and this does not include the benthic temperature. However, we argue that that the volume integrated global ocean temperature is dominated by the slower time scales of the deep ocean and hence is a good metric of ocean equilibrium, even for deep water. We have also examined the temperatures timeseries at 2731m (the deepest layer for which we have time series data) and it shows the same trends (see supplementary figures). We have added text at line 240.

We consider the volume integrated temperature because it includes all aspects of the ocean. However, it is dominated by the deep ocean trends and is near identical to the trends at 2731 (the lowest level that we have archived for the whole simulation).

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. We also point out that even the shortest simulation of 2000 years is longer than PMIP protocols.

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. The resulting time series of volume integrated global, annual mean ocean temperatures are shown in fig. 3. The supplementary figures also include this for each simulation, as well as the trends at 2731m.

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 updated the text (from line 176) but have to also note that the full details are commercially confidential. In practice, it should be largely viewed simply as a what-if sensitivity test.

Secondly, the smoother curve was heavily influenced by a previous (commercially confidential) sparser sequence of simulations using non-public paleogeographic reconstructions. The resulting

simulations were generally in good agreement with terrestrial proxy datasets (Harris et al., 2017). Specifically, using proprietary (commercially confidential) paleogeographies, we have performed multiple simulations at different CO₂ values for several stages across the last 440 million years and tested the resulting climate against commercially confidential proxy data (Harris et al., 2017). We then selected the CO₂ that best matched the data. For the current simulations, we linearly interpolated these CO₂ values to every stage. The resulting CO₂ curve looks like a heavily smoothed version of the Foster curve and is within the (large) envelope of CO₂ reconstructions. The first-order shapes of the two curves are similar, though they are very different for some time periods (e.g. Triassic and Jurassic). In practice, both curves should be considered an approximation to the actual evolution of CO₂ through time, which remains uncertain.

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. Moreover when studies use deep ocean temperatures to estimate global surface temperatures, they always assume linearity. We have added a line:

Although we could have used a non-linear fit, when we examine the residuals from the linear fit, they are near Gaussian suggesting that a linear fit is appropriate analysis. Moreover, all previous use of benthic temperatures as a global mean surface temperature proxy are based on linear relationship, as far as we are aware.

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 performance of the dynamic vegetation model is also described in (Valdes et al., 2017) but the deep ocean temperatures are not described in that paper. We therefore include a comparison to present day observed deep ocean temperatures 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 mean change in salinity has relatively little impact on the ocean circulation due to the near-linearity of the equation of state over much of the range of modelled salinities and temperatures.

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:

Simulations 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.

Line 230 – Impress number of simulations in near-equilibrium, but the values that define equilibrium seem a bit arbitrary.

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.

Line 265 – Why only below 1000 m? Uncertainty in shelf reconstructions? Also, there is a discrepancy with a figure text stating below 300 m.

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 .

Line 266 – I don't fully understand the reasoning behind using 2000 m temperatures. Please explain further. Is it to suggest that other records from 2000 m can tell us about the global benthic temperature? Is it to suggest that the model does not need to reach deep ocean equilibrium?

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 and hence there are strong gradients purely because of the depth variations. However, following the comment of the reviewer I have increased the depth to 2731m as it probably is more representative of the average from cores. It would be great for a data person to go over the existing compilations to better reflect the depth of the cores, but this would be a massive effort since the compilations have ~10000 data points and currently do not always record depth. However, for all diagnostics, it should be noted that the difference between the bottom temperatures and the 2116 or 2731m data were small. We have updated the text:

The observed benthic data are collected from a range of depths and are rarely at the very deepest levels (e.g. the new cores in (Friedrich et al., 2011) range from current water depths ranging from 1899m to 3192m). Furthermore large data compilations rarely include the depth profiles and thus effectively assume that any differences between basins and through time are entirely due to climate change and not to changes in depth. Hence throughout the rest of the paper we frequently use the modelled 2731m temperatures as a surrogate for the true benthic temperature. This is a pragmatic definition because the area of deep ocean reduces rapidly (e.g. there is typically only 50% of the globe deeper than 3300m). To evaluate whether this procedure gives a reasonable result, we also calculated the global average temperature at model layer 2731m. This is shown by the dashed line in figure 4a. In general, the agreement between model bottom water temperatures and 2731km temperatures is very good. The standard deviation is 0.7°C, and the maximum difference is 1.4°C. Compared to the overall variability, this is a relatively small difference and shows that it is reasonable to assume that the deep ocean has weak vertical gradients.

Line 282 – You say paleosols and stomata underestimate CO₂. Please provide citations. What CO₂ reconstructions does this leave for deep time?

I have modified this line. It probably isn't correct for paleosols and is difficult to find a reference for stomata (even though it is often discussed in workshops).

For these periods, the dominant source of CO₂ values for these periods is from paleosols (fig.1) and thus we are reliant on one proxy methodology. Unfortunately, the alternative CO₂ reconstructions of (Witkowski et al., 2018) have a data gap during this period.

Line 291 – Many other citations for model-proxy discrepancy in the Miocene (e.g. Goldner et al., 2014)

I have added Goldner (and also Krapp and Knorr). I selected You et al because I think it was the first to show discrepancy. I was not attempting a comprehensive review of the Miocene modelling.

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 $\delta^{18}\text{O}$ (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 we were surprised too, but I think it is the complexity of the problem. The mixed layer depth varies (spatially, seasonally and in duration) 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. Figure 6 already shows the different time periods in different colours.

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

If you mean “how does the high latitude respond to changing CO_2 etc”, then 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 CO_2 ?

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 shown in 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 without carefully controlled sensitivity simulations, it is difficult to quantify the relative importance.

Figures –

NOTE that the current figures are gif files but all are available with higher quality pdf files.

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 discussed in Farnsworth et al 2019 and in the surface polar amplification section (now section 3.4). We don't have any further insights.

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. Also added for each simulation in the supplementary text.

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 therefore removed this and pointed the reader to the Cramwinkel paper.

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. SSTs will be -2C below ice. Both lines added.

Figure 7 – Reason for the large change in average ocean depth at ~240 Ma?

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 paleogeographies.

Figure 10 – Dashed lines are a bit difficult to distinguish.

Improved