

Response to comments from reviewers

“Data-constrained assessment of ocean circulation changes since the middle Miocene in an Earth system model” by K.A.Crichton et al

An Overview of changes we propose to the manuscript based on the reviewers comments:

- Add simulations of a closed CAS from 10Ma and determine whether the model can better reproduce the data (as we have suggested in the manuscript). (Indeed, we have already created alternative continental configurations and re-run the pCO₂ vs. FwF ensembles for each of these.) We have already run statistics for benthic δ¹³C for a closed CAS between 10Ma to present, and model skill scores are higher than for the open CAS cases. We will perform the full analysis for temperature and δ¹³C on this new ensemble as for the one in the original manuscript. We will add to our discussion using the results of this new model ensemble.
- Better illustrate the CAS configurations that we apply.
- Add a specific discussion on cGENIEs fixed climate sensitivity and its implication for pCO₂ levels.
- Add ocean cross section plots for temperature and δ¹³C distribution as figure 14.
- Add a model sensitivity test of orbital configuration in supplementary material.
- Add to the supplementary material a detailed comparison of the ‘standard’ cGENIE climatology and ocean circulation vs. cGENIE configured following the late Holocene (‘modern’) HadCM3 simulation.
- Add some more information on sustained oscillations to supplementary material.
- Add a brief overview of ocean carbon storage changes in our ensembles in the supplementary material.
- Make improvements and revisions throughout the main text as recommended by reviewers (itemised below in specific response to their comments).

Reviewer 1

CAS configuration:

“As most of the discussion is CAS-oriented, authors should probably provide a few lines about CAS configuration changes during the Miocene and effect on the global circulation in the introduction using existing literature”

We will add more background from previous studies on CAS and its impact on circulation and climate in the introduction. We can and will also now link this discussion to the new alternative model continental configurations of CAS open vs. closed.

“As authors refers a lot to CAS closure, they should also probably provide as supplementary information a table that contains at least the mean depth of CAS in the different settings they used.”

We will add information on the depth and width of the model CAS in the supplementary material. Specifically – although Figure 3 shows the global bathymetry of each of the model continental configurations, we will add regional blow-up maps to better show the CAS bathymetry in model

detail. To this, we will add surface and sub-surface ocean velocity maps to illustrate the model simulated circulation through the CAS (and contrast with previous GCM-based modelling studies of this time interval).

"The paper would obviously benefit from additional sensitivity tests on CAS closure, especially for younger time slices (e.g. 2.5 and 4.5 Ma), as it seems, from what the authors states, that this is the missing point to be able to improve model-data fit."

We will carry out and present results from further simulations with a closed CAS (from 10Ma) to support our conclusions regarding CAS closure and $\delta^{13}\text{C}$ benthic patterns (and indeed have already completed the cGENIE ensemble simulations and are processing the model vs. data statistics). We will additionally include an analysis of the modern continental configuration, but with an open CAS, as a test of the data constraints in distinguishing an open vs. closed CAS.

Discussion:

"Moreover, as suggested by the title, the discussion should probably focuses on oceanic circulation; however in the present configuration, about half of this section is dedicated to GhGs forcing."

Rather than change the title, we will extend the discussion about ocean circulation, particularly in light of the new model ensembles and analysis with CAS open and the additional of new analysis of circulation flow through the CAS. We will also make it clearer at the start of the manuscript why we vary GHG forcing (as well as fresh water flux adjustments) and how this relates to constraining ocean circulation.

"As the section 4.1 is written now, it looks more like a list of previous Miocene studies than a real discussion about which may be the reasons why the CO₂ forcing required in this study to improve model-data fit does not fit with proxy-based CO₂ estimates."

In revising the manuscript, we will pay particular attention to this section. We will provide a new figure contrasting model-generated pCO₂ estimates with currently available proxies (as requested elsewhere), but as per the clarifications we will provide regarding 'why vary GHGs? if the study is focussing on ocean circulation' (the point above) we will make clear that this is a study designed to constrain surface climate not GHGs. Varying radiative forcing in the model is a means-to-an-end in firstly modulating the large-scale circulation of the ocean and hence benthic data patterns, but also in adjusting the mean benthic temperature and hence enabling us to employ more involved model-data stats such as the Model Skill Score rather than just e.g. the correlation coefficient.

"Also, methane (as well as gas hydrates) probably have a role to play but given all the issues associated with the absence of real atmospheric dynamics and vegetation feedbacks in the model, I am not sure it can by itself explain the discrepancy between the model and proxy-derived SST, while being forced with reasonable amount of CO₂ in the oldest-slices."

We do not propose that all the discrepancy is due to a lack of methane in the model (and acknowledge the limitations of our model that has no land surface representation). We will clarify this in the text. However, methane is a possible source of any discrepancy between our equivalent CO₂ and the actual Miocene CO₂ (which is not necessarily the proxy-data indicated value), so we consider it an important factor. We will streamline the text on methane to ensure its role is not too over-emphasised with respect to other sources of model eq.CO₂-realCO₂ mismatch (e.g. vegetation interactions etc.). We will also add new discussion about (assumed model) climate sensitivity and

the implications for model-required pCO₂ values, as in fact the proxy temperature constraints are arguably on radiative forcing in the model rather than pCO₂ per se.

"It is not clear if the biological pump is represented in the model configuration used for this study (ie. Is the biological pump module activated ?). I think this should be written more clearly"

We will clarify this in the text by adding a paragraph to summarise the model set up for ocean carbon cycling.

"In this study, the orbital parameters are kept as modern. However, at the time the proxies were generated it is likely that the orbital configuration was different, which can explain also part of discrepancy between proxy and model. I would have liked to see this point being discussed or at least mentioned in the discussion."

We will provide the results of a model sensitivity test of changing the orbital configuration. We did take account of orbital configuration for the selection of benthic data, by aiming for a mean (or representative) value in a window of +1 million years. We will add this to the text in the discussion, as well as in discussion of the surface temperature data, in terms of orbital configuration, that we obtained from other studies.

"Authors should also be careful to separate CO₂ forcing as it is in cGENIE (the one the authors refer to as equivalent CO₂ forcing... My advice would be to use "eq. CO₂ forcing" when the authors refer to the parameter used as a forcing in cGENIE."

In the text we will ensure that whenever we talk about the model forcing we will use eq.CO₂ to avoid confusion.

Other comments

We will apply all suggestions made in this section.

Technical corrections

We will apply all technical corrections except this one:

"Fig. 7; Fig. 11 Top - Please keep homogeneous min/max boundaries for the shaded color-scale within the different panels in the same figure. It make it easier to visually compare one time-slice to another. It also highlight changes in the simulated 13C"

We will apply the homogenous colour scale for figure 7, but not for figure 11. The δ¹³C global value changes during the study period, by ~1‰ according to our figure 13 for the atmosphere (well mixed) value. Keeping the colour bar the same for all timeslices in this figure 11 would mean any signal (from atmosphere – ocean differences) would be entirely dominated by this δ¹³C global value. The current colour scale in figure 11 shows the relationship between the atmosphere and the benthic δ¹³C value, where in some timeslices the benthic value is closer or further away from the atmosphere (or surface ocean value). We will clarify this in the text.

Reviewer 2

"1) I wonder how important paleogeography is to the response. The discussion contains speculation about the importance of gateways but I think this topic can easily be explored in greater detail. It would be interesting to compare the proxies from all time periods using only the Holocene configurations. Does paleogeography actually improve the model-proxy agreement? The authors already have all the results necessary for this comparison."

As per in reply to Reviewer 1 -- we will (and indeed have done now) perform CAS closed simulations for timeslices from 10Ma to test our original conclusion (that $\delta^{13}\text{C}$ data can be better explained with a closed CAS from 10Ma). To the specific point of performing model-data comparison using only the Holocene configuration – instead we propose the other way around – using the new open CAS Holocene configuration we have also generated, and contrasting with the Holocene data, the point being to test whether the most complete (Holocene) data set can distinguish open from closed CAS (when we know the answer for modern is ‘closed’). This point is additionally addressed via the new closed CAS 2.5 through 10.0 Ma simulations.

"2) I would like to see how the biases in HadCM3 are translated into cGENIE (muffingen) either by a comparison of the Holocene (HadCM3 forcing) and default cGENIE (observation forcing) simulations or by a comparison of the Holocene ocean circulation in cGENIE and HadCM3. This would help determine if the biases come from HadCM3 or cGENIE. A few sentences on the biases in cGENIE's ocean circulation would also be helpful."

We will include in the supplementary material a formal comparison of the HadCM3 forced vs the original observations-forced configuration of cGENIE, including an analysis of ocean circulation patterns and climatology.

"3) In the paper, the FwF changes are usually discussed with respect to changes in North Atlantic deep water. However, do most of the benthic proxy sites record a North Atlantic signal or an Antarctic signal? I realize that the changes are related, but talking about the Antarctic response might be more direct in some instances."

We will include discussion of any possible Antarctic signal with respect to the benthic data in the text.

Other comments

We will implement all comments suggested here. Some further explanation for selected comment is as follows:

"Line 111 – Are the reconstructions on paleolocations.org the same as the reconstructions used to make the maps in the HadCM3 experiments?"

They are not the same but we will provide error estimates in supplementary material between the two reconstructions (the differences are fairly small, we have now done this comparison). We found an error in one paleolocation in the datatables (site 1120), we will correct this in a revised version will re-run the statistical analysis on the full ensemble, but do not expect any significant change in the results of the statistical analyses (and certainly no change in our main conclusions).

"Line 188 – Why was this CO₂ coefficient chosen?"

This is the value predicted in the original carbon cycle tuning of Ridgwell et al. [2007] using the scaling from Wanninkhof [1992] in conjunction with climatological winds. In moving to a GCM-derived wind field, the spatial and temporal averaging going into the wind product is very different, meaning that one would expect the scaling between the wind speed product and gas exchange rate to change. We determine this new (GCM-based) tuning on the basis of retaining approximately the same global annual mean gas transfer coefficient for CO₂ as before. We will make this much clearer in the revision, and likely expand on this as part of the new discussion regarding climatology vs. GCM-derived cGENIE configurations (addressing previous points).

"Line 193 - Over how many years are the proxies averaged? I assumed the proxy averaging interval was long enough that orbital variability does not matter much, but based on the supplemental, it seems I was incorrect. I recommend averaging proxy records over 100 kyr or more to reduce the likelihood of capturing high frequency variability that is not simulated with the model. Also, might the use of present-day boreal summer near aphelion skew the results towards particular solutions?"

The data points were selected to be broadly representative of the pattern in the δ¹⁸O or δ¹³C over a window of +1 million years. This was not done by averaging over a certain interval but by visual inspection of the timeseries about the target age. The value selected was the mid-range of any evident glacial-interglacial cycling. Some of the data has far lower resolution than that would permit this method, so a window of +1 million years was inspected to determine the "representativeness" of a data value. This approach was taken due to the large uncertainty in age models from sediment core data, and in many datasets being far lower resolution than orbital cycles scales (e.g. less than one datapoint every 50,000 years for example). So in essence, we have taken account of orbital variability (where we can). To address the question of orbital configuration, we will add a model sensitivity test of changing orbital configuration to the supplementary material.

"Line 233 – The variability is so small that there is no benefit to averaging? I find this surprising."

The simplified 2D nature of the atmospheric component, in lacking interactive winds, leads to no interannual variability being present in the model. Hence at steady state, every year is effectively 'the same'. We will clarify this in the main text.

"Line 239 – I agree this is extremely interesting. Please list the simulations that show this behavior."

We are happy to add an illustration of this information to supplementary material.

"Line 243 – For oscillations that are longer (~4 kyr), how do you know they are persistent with only a 10 kyr long simulation? Please plot an example of these oscillations in the supplement? If you have to average over 4 kyr, are you in equilibrium over these 4 kyr?"

We see how our original text is a little misleading, there are actually no oscillations in any simulation longer than about 4.5 kyr. We will clarify the specific range we have determined. Will also provide an example of a 50 kyr long simulation in SI to illustrate the long-term persistence and stability of oscillations (also further addressing the previous request for more information).

"Line 299 – Can you modify these assumptions within uncertainty to improve model-proxy agreement?"

We will add additional short explanation as to the impact of these assumptions in the text, where it may have an impact on global ice volume at 15 Ma. However, a more extensive analysis of the impact of the uncertainty in global ice volume is probably a step too far. Temperature scales linearly with

changes in global ice volume (in the paleotemperature calculation that we apply), and as we don't have a strong error estimate for the global ice volume it is difficult to determine the impact on our results (except to say as we have that it has an impact in a certain direction).

"Line 326 – How do you determine the best fit value here?"

We only select values that we have tested (so, no intermediate values between 280ppm and 400ppm). A certain amount of the best-fit selection is a question of judgment, but we have tried to defend our selections in the text. In this case, 280ppm or 400ppm can be considered a good fit, but comparing to the Holocene range, a higher CO₂ (than Holocene) seems to be suggested. In figure 8 we note these high uncertainties by plotting a range of values (shaded grey) that also show some agreement with our data constraints.

"Line 487 – Again, you should be able to test the role of the CAS."

We will include the results of just such a test using our new series of alternative closed CAS simulation.

"Line 562 – Why is the fit so poor >70°?"

The high southern latitudes are far more effected by the polar front in the value of water δ¹⁸O, where sea-ice plays a strong role. Other latitudes follow the more simple salinity relationship. We will add some explanation to the text.

"Figure 8 – It would be helpful to plot CO₂ compilations against your best estimate (e.g. Foster et al., 2017; Berner and Kothavala, 2001)."

The CO₂ forcing that we find is an equivalent CO₂, so plotting it together with data-estimated CO₂ values is probably misleading. However, we can add a Figure of recent pCO₂ estimates, but in addition to plotting the cGENIE model-required pCO₂ values, we will also plot (i) model- required pCO₂ values corrected for potentially elevated atmospheric pCH₄, and (ii) model- required pCO₂ values transformed using a different climate sensitivity assumption – i.e. what the model and temperature data actually constrains is a radiative forcing, which in cGENIE equates to a particular pCO₂ value via the assumed climate sensitivity, hence assuming e.g. a higher climate sensitivity value as per recent fully coupled GCM results, would result in a lower required pCO₂ estimate for cGENIE. Note that the fixed/imposed and non state-dependent climate sensitivity in cGENIE allows us to simply re-calculate pCO₂ and not re-run the model ensembles.

"Figure 11 – Use a single color contour bar.

Figure 14 - Similar cross section plots of d13C and temperature would be very helpful."

As with reviewer 1 comment, we will not change the colour bars for figure 11. We will clarify in the text the value of keeping the colour bars on this sliding scale. We can add cross sections plots for temperature and d13C and will keep all colour bars on the same scale, which will address the point about comparing d13C absolute values.

Reviewer 3

"But the presentation here often just says "X constrains Y much more than Z does" without taking the step of diagnosing how the authors know that. A little more explanation would let the non-modeler get much more out of this analysis."

We will go through the text to identify where this occurs and improve the explanations.

Model set up:

"- can the authors assess the effect the lack of interactive clouds has on the sensitivity of surface temperatures to CO₂?"

As per our reply to Reviewer 2 and questions about the model-required pCO₂ values, we will add some discussion about the (fixed) sensitivity in cGENIE and the implications for the diagnosed pCO₂ values, including e.g. the absence of cloud feedbacks in the model that might otherwise lead to state-dependence of climate sensitivity.

"– Does the model allow for Albedo- vegetation feedbacks (I assume not)? As above, does this mean it will tend to underestimate Earth System Sensitivity to CO₂?"

According to some studies (e.g. Bradshaw et al 2015) yes, without vegetation feedbacks the model may underestimate Earth system sensitivity to CO₂. We discuss this in the discussion section, but we will clarify that by adding “we discuss climate sensitivity of cGENIE and the implications for our results in the discussion” in the model description.

"- nGCM input generated with CO₂ @ 400 ppmv"

We are happy to clarify and add discussion to the text in respect to the use of a series of GCM simulation assuming a single specific pCO₂ value to provide climatological surface boundary conditions, whilst significantly varying pCO₂ in the cGENIE model ensembles. This is related again to the fixed/imposed climate sensitivity in cGENIE, where the pCO₂ levels should be seen as a radiative forcing, this point from reviewer 3 can be addressed by discussion (and new figures) of the impact of methane and different climate sensitivity (as in response to reviewer 2) on the pCO₂ imposed.

"Lines 175-182 geographic/bathymetric choices such as not including Med, Greenland-NA connection, Bering Sea open/closed: can the authors provide the reader with a sense of whether these affect model results significantly"

Previous studies have suggested that these particular sea-ways are less important to global circulation changes in the Miocene-to-present than the Drake Passage and CAS. As such, we do not quantify specifically the effect of these seaways on the result of our analysis except to assume they are less important (based on other studies) than the Drake Passage and Central American Seaway. We will add a simulation of a closed CAS from 10Ma, so the CAS is in fact the only seaway we can analyse for its effect on circulation (after that closed CAS simulation is complete). Currently, we can compare the model output at different timeslices, but for the same forcings, to have an overall view of the effect of paleogeography (e.g. figure 5) but we can't attribute specific amounts of this effect to a particular seaway (this is outside the scope of this study).

We will explore much more clearly CAS bathymetry and bathymetry differences, and the consequences of this with the benefit of the new closed CAS ensemble.

"- Specify model sensitivity to CO₂? The text specifies "which is as the present day" without giving a value?"

It is approximate 3°C for a doubling of pCO₂. We will add this value, and state in the model description that we will further discuss climate sensitivity in the discussion.

"Emphasis on bathymetry and closing/opening of seaways on deep ocean properties: is it justified in comparison to more subtle modulators such as sea ice, ridge bathymetry, small changes in T and S? I'm thinking of ideas such as put forth for Ferrari et al. to account for glacial storage of CO₂ that rely on mixing or lack thereof of the North Atlantic and Southern Ocean deep circulation loops"

There is large uncertainty as to the CO₂ forcing and circulation states (and paleogeography) in this time period and we consider these to be the first order controllers on the global climate for the purposes of this study. Glacial storage of carbon is also an area of study in which uncertainties exist, but large scale paleogeography and CO₂ levels were (essentially) knowns for that time period (unlike ours).

We are aiming to create plausible “mid-range” (i.e. not extreme glacial or peak interglacial) climate and circulation states for our timeslices since the middle Miocene. The model determines sea-ice extent internally. In this sense, sea-ice is a result of climate state (not the inverse) for our study, but of course interacts with ocean physics and the carbon cycle. Ridge bathymetry can be seen as a smaller scale effect than the changes in bathymetry we consider that result in e.g. open or closed seaways for the Miocene period. Small changes in T and S are determined by our model physics: T we control with CO₂ forcing and S is (to an extent) controlled by our flux correction variable for the N. Atlantic. We consider that CO₂ forcing and circulation driven by that forcing and by paleogeography are much more significant controllers on global climate than those listed by the reviewer.

What we can easily do to aid here, is to add (to SI) an analysis of carbon storage in the ocean, which then for a single assumed value of pCO₂, provides the explicit information regarding the impact of changing circulation on stored carbon, and between ensembles provides information about how continental configuration (and circulation) impacts stored carbon. We note that although cGENIE incorporates a variety of e.g. preformed tracers, they were not included in the model configuration used in the ensembles. Adding then and then re-running and analysing the ensembles would admittedly be super interesting, but now way outside the scope of the present paper.

"I'm not sure why the authors have an open CAS for all except the Holocene time slice, since they note how circulation between the deep North Atlantic and Pacific affect model fits of carbon isotope gradients and North Atlantic salinity"

The CAS is open for all but the Holocene timeslice due to the method we apply for regridding the HadCM3 model, which is of finer resolution than cGENIE. We will add a new set of simulations with a closed CAS from 10Ma to test our conclusions (see replies to previous points) and also make clearer how the first set of continental configurations came to be (in terms of the re-gridding algorithm).

"Circulation and deep ocean CO₂ inventory- is this examined explicitly? We know from Pleistocene CO₂ cycles that the storage of carbon in the Southern Ocean can give a 80-90 ppm CO₂ effect. Some theories (e.g. R. Ferrari et al. 2014) hypothesize that the connection between the North Atlantic and Southern ocean loops is critical to whether CO₂ is stored in the deep ocean or vented to the

atmosphere. Can the authors comment on how the Southern Ocean CO₂ pump works in cGENIE and if the model is capable of capturing changes such as proposed by Ferrari on carbon storage?"

See reply above – yes we can add a brief analysis of ocean carbon storage and how continental configuration and circulation patterns impact this. Associated with this we can also include a very brief comparison with ocean LGM carbon storage, but limited to avoid straying too far into off-topic glacial/interglacial questions.

"what is the balance of alkenone vs TEX86 data? Does the balance change between time slices? Which TEX86 index was used?"

We will add a note stating which referenced surface temperature set is from which proxy in the datable.

"As the deep ocean has extremely small salinity variations, it's not clear why a salinity correction is needed. The deep ocean salinity will tend to be pretty homogeneous, because it is filled from only a handful of sources. Furthermore, the deep ocean salinity is constrained to be very close to the average salinity of the entire ocean, since only the thin skin layer of the ocean exchanges water with the atmosphere. The only way to make large changes in average salinity on the timescale considered here is to remove/add water by forming/ melting ice. The correction for the ice volume effect on isotopic composition, is potentially large and not very well constrained from the Pliocene and earlier, as the volume and the isotopic composition of polar ice aren't well known... The salinity of the global deep ocean is constrained to be the average salinity of the ocean (because deep water masses have only one or two sources and must have nearly the same densities throughout the global deep ocean); likewise for the d18O value of the deep ocean water. Therefore, I think the language above should be clarified so the "the d18O signal recorded" is rewritten as "the d18O signal recorded in the North Atlantic: : :" or similar. I take it from the topic sentence of the paragraph that the authors are looking at local patterns but the reader may mistake the intent of the later sentences."

Supplementary material figure A2 of deep ocean $\delta^{18}\text{O}_{\text{sw}}$ demonstrates the need to correct for $\delta^{18}\text{O}_{\text{sw}}$ spatial variability in the paleotemperature calculation (rather than assuming a global fixed value). Supplementary figure A1 demonstrates the effect of this spatially variable $\delta^{18}\text{O}_{\text{sw}}$ on the calculated temperature. Salinity strongly affects $\delta^{18}\text{O}_{\text{sw}}$, and this salinity value comes not only from global ice volume (where ice is fresh, so the ocean is more saline for large ice volume) but from everyday effects of evaporation and precipitation at the ocean surface, and from these water masses circulations, which transport more or less saline water to the deep; the "global ocean conveyer" is the thermohaline circulation (haline referring to salinity).

The cGENIE model traces salinity through evaporation, precipitation and circulation, so we use cGENIE salinity to correct for local $\delta^{18}\text{O}_{\text{sw}}$. As salinity affects $\delta^{18}\text{O}_{\text{sw}}$ (again, due to evaporation and precipitation because less energy is required to evaporate ^{16}O than ^{18}O , and less energy is required to precipitate ^{18}O than ^{16}O) and these more or less salty waters are differently distributed (by ocean circulation), the deep ocean is not homogenous in $\delta^{18}\text{O}_{\text{sw}}$ and we need to correct for this non-homogenous water $\delta^{18}\text{O}_{\text{sw}}$ *at every location* to get temperature from foram shells $\delta^{18}\text{O}$ (using the Marchitto et al 2014 paleotemperature calculation). As we think the thermohaline circulation with the Atlantic overturning leg established in its present form since the middle Miocene, this correction for local $\delta^{18}\text{O}_{\text{sw}}$ becomes even more important for our study. By applying this salinity correction to $\delta^{18}\text{O}_{\text{sw}}$ some benthic temperature estimates change by up to 3°C, which is clearly a significant amount. The correction (and need for it) is described in Appendix A. We will go through the text to ensure it is well sign posted and clear.

"I'd like to see more discussion of how the atmospheric d13C was derived- the presentation here is too short to be useful to someone not intricately familiar with cGENIE."

The ocean $\delta^{13}\text{C}$ scales linearly with atmospheric $\delta^{13}\text{CO}_2$. We simply used the difference between the data mean benthic d13C values and the mean model benthic $\delta^{13}\text{C}$ values (which were all forced with d13CO₂ of -6.5) and used the offset between these to determine the $\delta^{13}\text{CO}_2$ – i.e. the $\delta^{13}\text{CO}_2$ value that would minimise the mean model-data benthic $\delta^{13}\text{C}$ offset. We will add a clarifying sentence to the existing text.

"Woodruff, F. and S. M. Savin (1989). "Miocene deepwater oceanography." Paleoceanography 4(1): 87-140."

We use data from this study in our dataset. In general, we did not want to rely on existing data-interpretation studies to interpret our model-data results, as we rely on the globally distributed datasets from many sources to constrain circulation (rather than specific early studies).

"I wondered if the d13C data base has a change in the proportion of Uvigerina vs Cibicidoides values over time? While we "correct" for isotope offsets between species, this may not always work."

We only use Planulina and Cibicidoides species to compile our dataset in order that we can apply the specific paleotemperature calculation of Marchitto et al 2014 (this is already specified in the text). We are aware that different species have different sensitivities to temperature for their shells $\delta^{18}\text{O}$, hence why we restrict the species.

"Finally, the style of the paper wavers a bit between assuming no previous background and a lot of background. I'm think this level of explanation is not needed for the target reader- it's too low-level (lines 89-92): "The isotope of carbon, carbon-13, has been used as a tracer for paleo ocean circulation for many years (Lynch-Stieglitz 2003). It is a stable isotope, heavier than carbon-12 and accounts for about 1% of all carbon on Earth. The ratio of carbon-13 90 to carbon-12, designated as "13C" – the divergence from a standard in units of parts-per-thousand (‰, can be estimated for paleo ocean waters by measuring the 13C in shells of foraminifera formed in those paleo water masses: :: Shells of dead foraminifera gradually accumulate on the ocean floor, thus providing a record of changes in water column chemistry over time at that location in the form of ocean sediments""

We will go through the manuscript to ensure a more consistent level of detail and assumed reader background.

We thank all reviewers for their comments.