

# Response to Referee 1

---

---

We are very grateful to the referee for the time and effort expended in reviewing our manuscript. We believe that these comments and suggestions have enabled us to improve the presentation insofar as its clarity is concerned.

The results presented here exhibit appreciable differences from previously published PI and PlioMIP1 experiments with CCSM4. The claim that this is due to the extended simulation length is not convincing as the surface temperature field appears to be in quasi-equilibrium after about 500 years (Fig. 5); extending the simulation beyond this point only adds about 0.5°C to the global annual temperature.

It is important to note that we have never claimed in our manuscript that our ability to capture PRISM3 SST anomalies much more accurately with our models than was possible with CCSM4 in the first phase of PlioMIP is only (or even primarily) due to the very long timescale over which the simulations have extended. In fact, we have made it clear that it is the nature of the new boundary conditions that have contributed primarily to the much improved data-model comparison that our work has achieved. However, the implementation of these new boundary conditions has required significantly longer integration times to reach conditions that are sufficiently close to equilibrium, as we argue in greater detail below in response to another of the reviewer's comments. Note specifically the following excerpts from the submitted text:

Abstract: “With the new boundary conditions, the CCSM4 model simulates a mid-Pliocene which is...”

Conclusion: “We find that the PRISM4 boundary conditions mandated in PlioMIP2 lead to greater warming in the mid-Pliocene”

It is also incorrect to assume that because the globally averaged SST might have reached quasi-equilibrium that this also implies that the SSTs have come to equilibrium regionally. Since, at year 500, the MOC continues to move towards its new equilibrium and the TOA energy imbalance continues to decrease, the model is still in a state of significant disequilibrium.

The authors are urged to (1) double-check their simulation setup, (2) run a PlioMIP1 simulation using the same setup as in Rosenbloom et al., (2013) to assess differences from previously published experiments with CCSM4 (data can be downloaded from the Earth System Grid: <https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm4.output.html>).

Given the significant difficulties insofar as the utility of model-data comparisons based upon the initial design of the PlioMIP effort is concerned, it is our opinion that nothing could possibly be gained by regressing to that initial design and wasting the very significant computational resources that this would entail. This is especially the case given the very marked improvement of the data-model inter-comparisons that have been achieved with the new experimental design. Our analyses are the one of the first to be produced based upon this improved design.

DR's simulations (if proven valid) may be interesting for the purpose of testing CCSM4 parameterizations, but they may not be appropriate for PlioMIP2. One major goal of PlioMIP2 is to evaluate skills in simulating mPWP climate by state-of-the-art Earth-system models. DR's PI simulation shows much worse model skills (too cold and too much sea ice) compared to the published benchmark CCSM4 PI simulation. It is confusing to use these simulations to serve the intercomparison purpose for CCSM4.

These comments are also ill-founded for reasons that will be further clarified in what follows. In fact, the pre-industrial control model employed in *Rosenbloom et al.* [2013], while it provided a reasonable representation of the climate of that era, suffered from significant flaws as a control model on the basis of which one might attempt to ascertain the deviation of mPWP climate from that of the pre-industrial era. In particular, the PI control employed by *Rosenbloom et al.* had both the “overflow” parameterization and the “tidal mixing” parameterizations turned on. Given that the new boundary conditions have significantly modified bathymetric depth throughout the global oceans and given that the land sea mask is also modified, it is expected that these modern climate-based parameterizations would be inappropriate for the mPWP on a priori grounds. We have therefor turned them off to produce a unique PI control with which our mPWP results can be compared, unencumbered by the inconsistency of the comparisons previously produced using CCSM4 in the context of the original PlioMIP exercise. This is further discussed in what follows.

Finally, there is a lack of information on the spin-up and diagnostics of DR's mPWP simulations, which hinders comprehensive evaluations of validity of these simulations. I am highlighting major differences from published CCSM PI runs here, using the part of DR simulations that do not include changes to ocean diapycnal diffusivity:

Before we individually address the specific differences between our CCSM4 PI and *Gent et al.* [2011] PI that the reviewer has referred to, it is necessary to make it clear that *Gent et al.* is but a single realization of the PI control run which was produced by the authors for a single specific configuration of the CCSM4 model (overflow and tidal mixing parameterizations turned on) and for a specific number of model years. It is entirely expected that different configurations of the same model, and different lengths of model integration will result in a set of climatologies that differ somewhat among them, while being broadly similar. There is no single PI simulation that can be proposed to be better than any other PI simulation because such a determination cannot be made for a time long before the start of modern observational era. What is actually important is that the configuration of the model employed for the PI control not contain parameterizations that are expected not to be applicable under mPWP conditions.

Here are some of the ways in which our PI simulation differs from *Gent et al.* [2011].

1. The *Gent et al.* simulation was integrated for a total of only 1,300 model years whereas our PI control run has been integrated for a total of 5,170 years which is almost four times longer than the “benchmark simulation”. There is no physical basis on which the climatology of two simulations differing by almost 4,000 years in model years should agree.

2. We share our own concern with that of the referee in regards to the importance of establishing a steady state control with minimal model drift (discussed further below), and it is for that reason that we have undertaken the very challenging task of simulating a  $1^\circ$  coupled atmosphere-ocean PI control for so many model years. This has resulted in a model that is much closer to equilibrium than the control of *Gent et al.*, which continued to lose heat at the top of the atmosphere (TOA) at a globally averaged rate of  $-0.147 \text{ W m}^{-2}$ . This is larger than the  $< |0.1| \text{ W m}^{-2}$  energy loss rate that *Gent et al.*, themselves consider desirable and larger than the  $0.11 \text{ W m}^{-2}$  energy balance that we get in our PI control  $E^{280}_P$ . Furthermore, the  $E^{280}$  model from which  $E^{280}_P$  was branched off reached an energy balance of  $0.02 \text{ W m}^{-2}$  implying an essentially perfect equilibrium state. Continued integration of  $E^{280}_P$  will lead to further reduction in the TOA energy imbalance.
3. The ocean model in *Gent et al.* CCSM4 PI simulation was configured with both (i) the overflow parametrization and (ii) tidal mixing scheme turned on. As we have discussed in our manuscript, and which we further discuss below in response to the reviewer's concerns, these two options of the ocean model are tuned to the present day conditions and as such should not be used in palaeo climate simulations [see *Vettoretti and Peltier [2013]*, *Peltier and Vettoretti [2014]* for comments on this point in connection with their work on the Dansgaard-Oeschger oscillation phenomenon] for epochs in which they are expected to be inappropriate on a priori grounds.
4. There are also some differences between the ocean mask in our PI control and that in *Gent et al.* See *Vettoretti and Peltier [2013]* for details of the differences. Of course the land-sea mask under mPWP conditions differs significantly from modern and this will have a strong impact upon the tidal regime on the basis of which a new tidal mixing parameterization, appropriate to that era, would have to be determined.

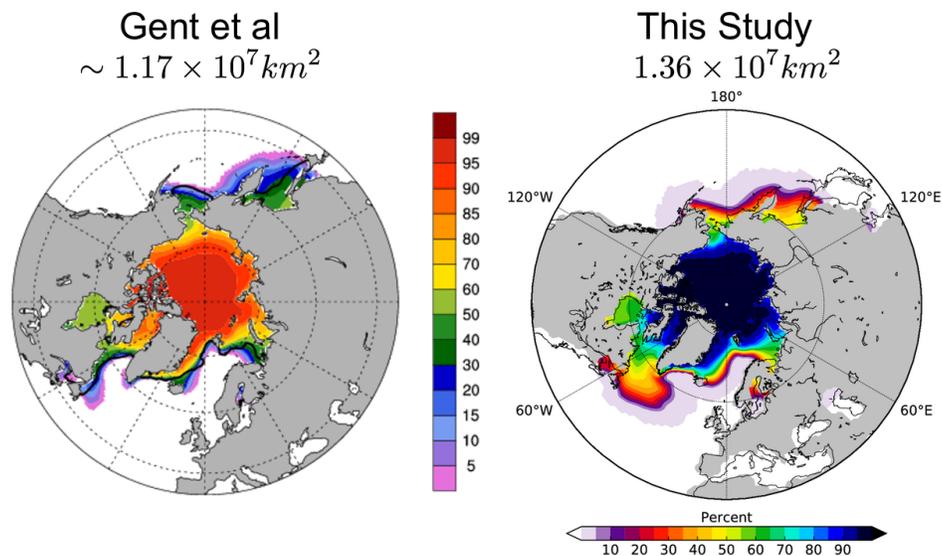
1. The global surface temperatures in DR's PI simulation is between  $12.5$  to  $13^\circ\text{C}$  (Fig. 5 and Table 2), which is  $1^\circ\text{C}$  colder than CCSM4 PI benchmark simulation ([http://www.cesm.ucar.edu/experiments/cesm1.0/diagnostics/b40.1850.track1.1deg.006/atm\\_863-892-obs/](http://www.cesm.ucar.edu/experiments/cesm1.0/diagnostics/b40.1850.track1.1deg.006/atm_863-892-obs/)); the benchmark simulation shows  $13^\circ\text{C}$  for 2-m air temperature and  $14^\circ\text{C}$  for surface temperature. Northern hemispheric sea ice in DR's PI case covers Labrador Sea and North Pacific at least seasonally (Fig. 15), which is way too extensive compared to the published PI benchmark simulation (*Gent et al.*, 2011). In *Gent et al.*, (2011), the CCSM4 PI benchmark has an AMOC strength  $> 24 \text{ Sv}$ . Shown in *Yeager and Danabasoglu (2012)* Figure 1b, turning off the overflow parameterization does not change the AMOC strength much, which is measured as the maximum overturning of the North Atlantic. However, in DR's PI simulation, the AMOC strength is only  $20 \text{ Sv}$  (Table 4.). Finally, the entire ocean in DR's PI simulation is colder than what is reported by *Gent et al (2011)*.

The minor differences that the reviewer has noted between our PI control and that by *Gent et al. [2011]* are entirely expected on the physical grounds that our PI control has been integrated for a period that is four times longer than that in *Gent et al.* and has both the overflow

parameterization and tidal mixing parameterization turned off. Therefore, the reported climatological features of our model represent a different (and more fully developed) equilibrium state than that in *Gent et al.*

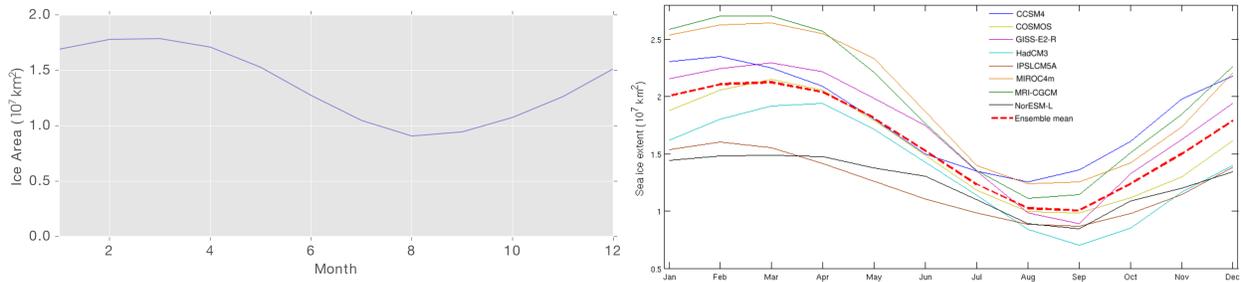
The temperature reported in Figure 5 of the manuscript is not 2 m air temperature, which is something we clarify below in response to another of the referee’s comments. This has also been clarified in the revised manuscript. The values in Table 2 are the 2 m air temperature which for our PI control model is 13°C, a value which is in very good agreement with the 2 m air temperature for the benchmark simulation (13°C according to the referee).

With regards to the sea ice in the Labrador Sea and the North Pacific, these differences are due to the different configurations of the ocean component, and the significantly different duration of the simulations, as discussed above. Furthermore, in the absence of any reliable observations regarding the extent and volume of PI sea ice there is no reason to assume that either the result of our simulation or that of *Gent et al.* is superior to the other. However, it does appear questionable as to why the margins of the mean annual arctic sea ice concentration in the *Gent et al.* PI control are as close to that which has been inferred on the basis of modern satellite observations (Figure 3 in *Gent et al.*, also reproduced below)? The equilibrium state of the 280 ppmv PI climate would be expected, on *a priori* grounds, to be characterized by more extensive Arctic sea ice formation than that which has been observed for the recent decades and this is what is predicted by our unique PI control model. The following Figure compares the mean annual arctic sea ice in our PI (left) and in *Gent et al.* (right). The sea ice coverage in our PI is only slightly larger (~ 16%) than *Gent et al.* ( $1.36 \times 10^7 \text{ km}^2$  in ours compared with  $\sim 1.17 \times 10^7 \text{ km}^2$  in *Gent et al.*).



Interestingly, the PI control that has been used by *Rosenbloom et al.* [2013] in PlioMIP (which is an extension of *Gent et al.*, in that the reported PI climatology is from the last 100 years of the 1,300 year *Gent et al.* control, whereas *Gent et al.* report on years 871-900 of that same simulation) has more arctic sea ice than our PI control. The Figure on the next page shows the monthly cycle of Arctic sea ice extent for our PI control (left), and the PI controls of all PlioMIP

models on right (from *Howell et al.* [2016]). We note two things (i) the large spread among the PlioMIP models for the PI sea ice seasonal cycle, particularly during the winter months, and (ii) our PI seasonal cycle follows the ensemble mean closely, especially during June-December. Annually averaged, the sea ice extent in our CCSM4 PI is  $1.36 \times 10^7 \text{ km}^2$  and that in *Rosenbloom et al.* [2013] is  $1.83 \times 10^7 \text{ km}^2$  (as reported in *Howell et al.* [2016]), further arguing against the referee’s assumption that there is too much sea ice in our PI control model. These comparisons leave no doubt as to the reasonableness of our PI sea ice cover.

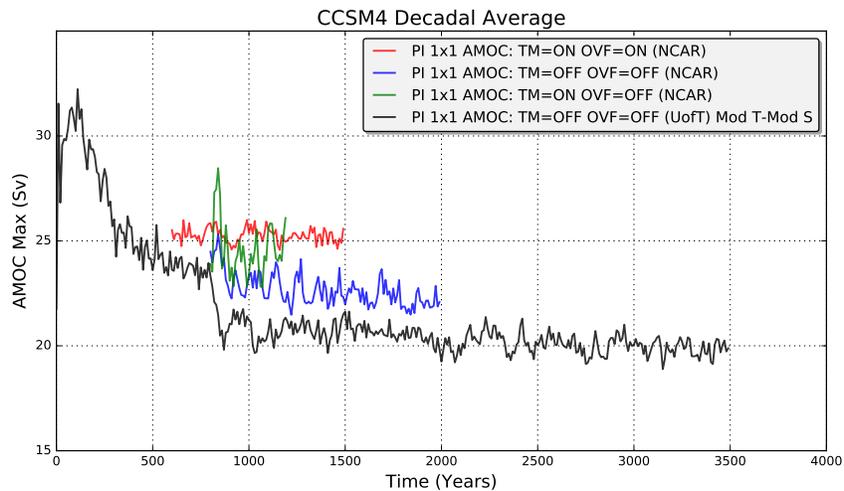


The AMOC strength that the reviewer is quoting is from the version of our PI control with the pelagic value of the diapycnal diffusivity. Our PI control with the more realistic POP1 profile of diapycnal diffusivity that is employed in the KPP parameterization is 21.5 Sv. Observational estimates for the strength of the AMOC are in the range of 17-18 Sv [*Cunningham et al.*, 2007, *McCarthy et al.*, 2015]. Therefore the 24 Sv estimate by *Gent et al.* [2011] is the problem, and our own estimates are larger than the observationally estimated range, although, compared with *Gent et al.* our results are closer to observations. The CCSM4 PI AMOC strength that was used for the intercomparison of AMOC among PlioMIP models was even larger at 25.7 Sv [*Zhang et al.*, 2013]. It seems that CCSM4 has a consistent bias towards predicting AMOC strengths that are too large. In the intercomparison by [*Zhang et al.*, 2013] the CCSM4 AMOC strength was an outlier among the eight models compared; five of the models predicted an AMOC in the range of 14-19 Sv.

With regards to the reviewer’s comment that elimination of the overflow parametrization does not change the strength of the AMOC “much”, we show on the figure on the next page, a preliminary result that turning off the overflow parametrization leads to a change of  $\sim 1$  Sv (comparing green and red curves). The red curve is the AMOC strength as predicted by the NCAR configuration for PI (i.e. that of *Gent et al.*), the blue curve is the same NCAR model which has tidal mixing and overflow parametrization turned off. Finally, the black curve shows the PI control from which our control simulations E<sup>280</sup> and E<sup>400</sup> are started. This speaks to the excellent equilibrium of the pre-existing control from which our PlioMIP controls are initiated.

The referee’s claims that the entire ocean in our PI is colder than *Gent et al.* is manifestly incorrect. The globally averaged ocean temperature reported for the PI in *Gent et al.* is 3.13°C, whereas the globally averaged temperature in our PI (E<sup>280</sup><sub>P</sub>) towards the end of the simulation is 3.1 – 3.2°C.

2. The climate sensitivity to a doubling of CO<sub>2</sub> is  $\sim 4^\circ\text{C}$  (based on simulations E400 and E280), while the estimate for CCSM4 is  $\sim 4^\circ\text{C}$ /per doubling of CO<sub>2</sub> (Bitz et al., 2011).



The reviewer is confusing the Equilibrium Climate Sensitivity (ECS) which is the subject of *Bitz et al.* [2012] with Earth System Sensitivity (ESS), which is the subject of our paper and which has been employed in the context of PlioMIP [*Haywood et al.*, 2013]. ECS (also sometimes known as the Charney Sensitivity) is defined as the change in the average surface temperature of the planet in response to a doubling of CO<sub>2</sub> concentrations, determined after a sufficient amount of time has passed to allow for the fast feedback components of the climate system such as water-vapor, aerosols, clouds, sea-ice albedo etc. to reach equilibrium. In contrast, a model's ESS is computed after allowing sufficient time for the slow feedback processes such as those associated with ocean circulation, and land ice and vegetation (the latter two only applicable if the climate model is configured with them) to reach equilibrium. Consequently, ESS is always larger than ECS. In palaeo-climate simulations where one's interest is in the equilibrium state it is the ESS that is of interest. For the original PlioMIP, the CCSM4 model's ESS was determined to be 3.51°C/2 × CO<sub>2</sub> [*Haywood et al.*, 2013] while the ECS of CCSM4 is 3.2°C/2 × CO<sub>2</sub> [*Bitz et al.*, 2012].

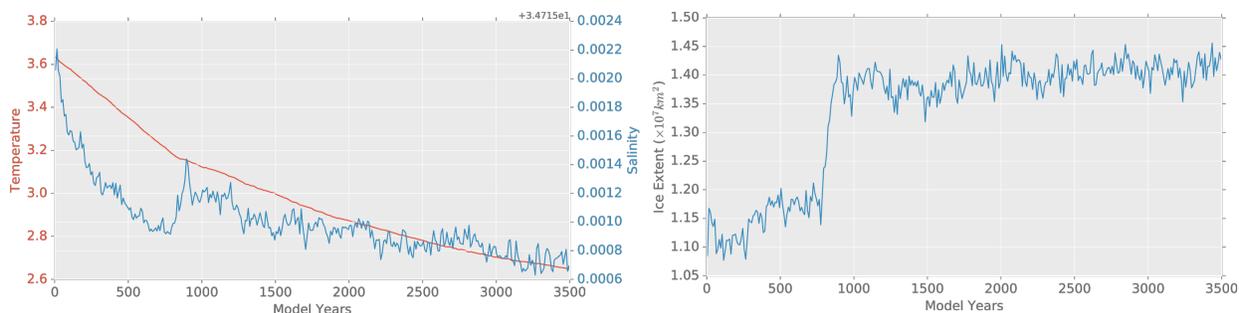
3. Comparing the zonal mean temperatures in DR's E400 to E280 case (the distance between red and green lines in Fig. 8), the 120 ppm CO<sub>2</sub> increase leads to ~ 7°C warming in Arctic region. In IPCC AR5 (Fig 12.11 in Collins et al., 2013), ~ 100 ppm CO<sub>2</sub> increase between 2081-2100 and 1986-2005 based on RCP2.6 simulations shows < 3 – 4°C Arctic warming.

The referee has here made the further error of confusing the temperature anomaly between two equilibrium climates with the temperature anomaly between two transient states. This error is related to the reviewer's mistake in confusing ESS and ECS which we have discussed above. The climates for years 2081-2100 and for years 1986-2005 are not in equilibrium, rather the climate in these IPCC models is continually forced by historical and RCP emission scenarios. In each of these two intervals the slow feedback components of the climate system haven't even come close to equilibrium, and one might even question as to how accurate it is to assume that the fast feedback processes might have reached equilibrium while the climate system continues

to be forced by time varying emissions. Therefore, the arctic warming in *Collins et al.*, [2013] is expected to be smaller than that which would be obtained when comparing two multi-millennia climate runs such as E<sup>280</sup> and E<sup>400</sup>.

More information is needed to evaluate DR's mPWP simulations: 1. Details about the pre-existing 3500 year spin up simulation should be presented. This run was used to initialize the E280 and E400 case. For example, what are the salinity trends in major ocean basins and what are the sea ice trends? how were the carbon and nitrogen pools initialized and spun up to equilibrium? Moreover, how does the simulated E400 climate state compare to present-day observations? The surface temperatures presented in Fig. 5 suggest that both E280 and E400 start from a very cold ocean state, which is an odd practice for simulating PI or present-day climate.

We refer the reviewer to *Vettoretti and Peltier* [2013] for a discussion of the pre-existing 3,500 year spin up. The ocean salinity and temperature trends and the sea-ice trend are shown below for the pre-existing spin-up. These analyses show that the spin-up simulation is in a steady-state equilibrium. The global ocean temperature trend over the last 500 years is only 0.01°C/century.



A comparison between the simulated E<sup>400</sup> climate state and present-day observations is not part of the manuscript because:

1. Such a comparison is not pertinent to the topic of the mid-Pliocene.
2. The E<sup>400</sup> simulation in PlioMIP2 is not designed to represent the present-day. It does not include present day aerosols, CFCs, emissions scenario, urban land units and agricultural land units. Its sole purpose is to serve as a 400 ppmv control.
3. The E<sup>400</sup> leads to an equilibrium climate and therefore there isn't much value in comparing it to the present day transient climate. The CMIP project is devoted to understanding model inter-comparisons for the present-day.

The E<sup>280</sup> and E<sup>400</sup> are branched from the pre-existing control run, as such the temperature evolution of the two models begins near the equilibrated ocean state of the 3,500 year long PI control. It is not clear to us why the reviewer was confused by this given that the reviewer seems to understand that the two simulations are branched from a well equilibrated PI run (which was started from *Levitus and Boyer* [1994], which we clarify in the next point). Nevertheless, we hope that our response has clarified this for the reviewer.

2. Why are the Pliocene simulations initialized from Levitus? This shows a very different provenance from the controls. Earth system sensitivity analyses should not be performed when the oceans are started so differently and not brought into any kind of equilibrium (clear ocean temperature trends are shown in Figure 4 for PlioMIP2 cases). Based on Fig. 4, difference in ocean temperature between  $E^{400}$  and  $Eoi^{400}$  could be an artificial result of different initialization.

The perpetually running control from which  $E^{280}$  and  $E^{400}$  control experiments were branched was also started from *Levitus and Boyer* [1994] [see *Vettoretti and Peltier* [2013]]. We realize that this information wasn't clear from our discussion of the control experiments in section 3.2 and we have made the changes in the revised version of the text needed to clarify this point. Therefore, all simulations we report upon have been initialized from *Levitus and Boyer* [1994].

3. The argument for vegetation mapping is weak (P8, Line 30-34). For example, for Megabiome 1, 80% PFT type 5 and 20% PFT type 7 could lead to different surface albedo from 20% PFT type 5 and 80% PFT type 7. More rigorous mapping method should be derived and applied to the PlioMIP2 simulations.

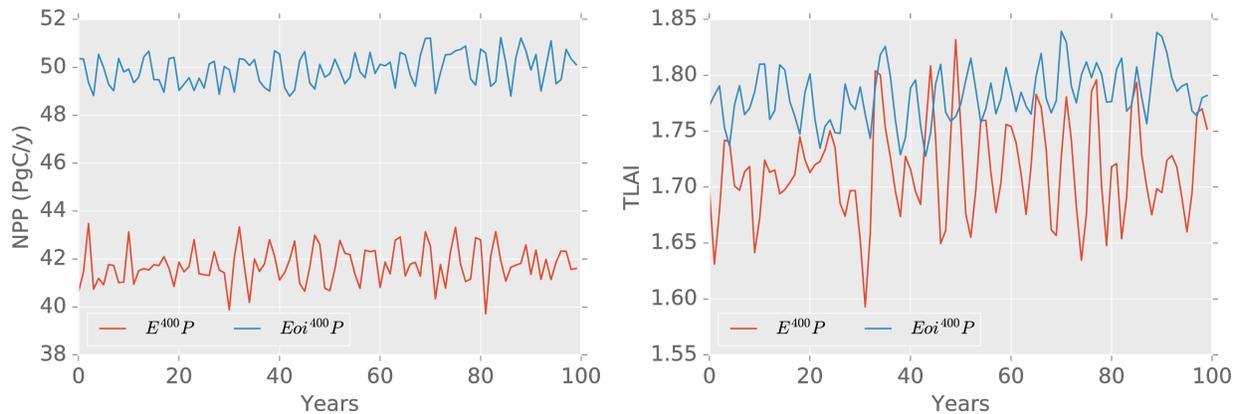
Paleo-vegetation reconstruction is a significant source of uncertainty that continues to plague all palaeo climate simulations including the simulations that have been published in the context of PlioMIP1 and those that are currently being carried out for PlioMIP2. In order to maintain consistency with PlioMIP1 simulations and the PlioMIP2 simulation using the MRI-CGCM2.3 model [*Kamae et al.*, 2016] we use the mid-Pliocene reconstruction by *Salzmann et al.* [2008]. The mapping from the mega-biomes to PFTs that we have employed is based on the relative projection of each of the mega-biomes onto PFTs for the modern day. Therefore, while there is a certain amount of latitude in the relative makeup of a mega biome in terms of PFTs, the example that the reviewer has used is unlikely because that situation would not constitute a reasonable projection of that mega biome onto PFTs. *Rosenbloom et al.* [2013] have used a similar methodology for initializing their vegetation cover, and although their choice of map was not made available as part of their model publication, we feel it is important for us to make our map available so that other groups might use it in their Pliocene simulations if they so choose.

4. The significant and widespread cooling in DJF over Canada and Eurasia in the  $Eoi400$  simulation as compared to the  $E400$  simulation needs to be explained (Figure 7). The cooling suggests a possible problem with the initialization of the land surface in the  $Eoi400$  simulation. Please show time series of net primary productivity and leaf area index. An example of Rosenbloom et al (2013) can be found here: <http://webext.cgd.ucar.edu/B1850/Pliocene/lmwg/ccsm2/b40.plio.FV1.003ext-b40.1850.track1.1deg.006/set1/set1.html>.

We have mentioned in our manuscript that the nature of cooling in DJF over Canada and Eurasia in the anomaly of the Pliocene simulation with 400 ppmv modern bears special consideration due to its potential similarity to observed cooling trends over recent decades [*Cohen et al.*, 2013,

Overland *et al.*, 2011, Sun *et al.*, 2016] and therefore a comprehensive examination is planned to be the subject of a further publication.

As per the request of the reviewer, we show in the next figure the time series of NPP and TLAI over the last hundred years of simulation for both the 400 ppmv Pliocene and the control. These results show a robust and well equilibrated vegetation and there is no sign of any problem with the initialization of the land surface in the Pliocene simulation.



5. Global, annual surface air temperature in the Eoi400 simulation is  $\sim 2^{\circ}\text{C}$  warmer than in the E400 simulation (Figure 5, Table 2). This is puzzling, even more so because  $\sim 1^{\circ}\text{C}$  of this surface temperature difference occurs at tropical latitudes (Figure 8) where the relatively small coverage and changes in vegetation would not be expected to produce that much warming. If this is indeed a result of the changes in the gateways then a figure supporting this contention needs to be included.

The reviewer asserts that the simulated anomalies are “puzzling” but has not articulated as to why? With regards to the temperature difference along the tropical latitudes, this is due to the significant expansion of warm pool waters throughout the tropics globally (manuscript Figures 10, 11), with which the atmosphere exchanges latent and sensible heat.

6. As noted above, DR’s PI simulation shows too much sea ice in the North Atlantic and Labrador Sea compared to the published benchmark CCSM4 PI simulation. As such, it strongly influences their surface air temperature anomalies in the North Atlantic and Labrador Sea, as shown in Figures 6a and 7a, leading to their conclusion that they better match the PRISM3 SST anomalies in these regions (Figure 12) than the simulation of Rosenbloom *et al.* In fact, their PlioMIP2 simulation, by itself, is colder than Rosenbloom *et al.*, (2013).

The reviewer’s comment that our simulation is colder than Rosenbloom *et al.* [2013] is puzzling considering that the volume-integrated global ocean temperature in our 400 ppmv Pliocene simulation is  $\sim 5.8^{\circ}\text{C}$  (manuscript figure 4), compared to  $\sim 4^{\circ}\text{C}$  in Rosenbloom *et al.* [2013] (Figure 2c). We point out that this is the second time that the reviewer has made

an incorrect statement regarding the comparison of ocean temperature between one of our simulations and one of the reviewer's preferred set of simulations. The first mistake was when the reviewer incorrectly claimed that our PI ocean is colder than *Gent et al.* [2011].

Because the reviewer has repeated the same comments with regards to our PI control simulation, we are obliged to address them again with the same arguments presented earlier:

1. The *Gent et al.* [2011] simulation was integrated for a total of only 1,300 model years whereas our PI control run has been integrated for a total of 5,170 years which is almost four times longer than the “benchmark simulation”. There can be nothing sacred concerning the *Gent et al.* PI run as a CCSM4 “benchmark” for use in the context of PliomIP. There is no physical basis on which the climatology of two simulations differing by almost 4,000 years in model years should agree.
2. Because our PI control is integrated for almost 4 times longer than that in *Gent et al.* [2011], it has resulted in a model that is much closer to equilibrium than the control of *Gent et al.*, which continued to lose heat at the top of the atmosphere (TOA) at a globally averaged rate of  $-0.147 \text{ W m}^{-2}$ . This is larger than the  $< |0.1| \text{ W m}^{-2}$  energy loss rate that *Gent et al.* consider desirable and larger than the  $0.11 \text{ W m}^{-2}$  energy imbalance that we get in our PI control  $E^{280p}$ . Furthermore, the  $E^{280}$  model from which  $E^{280p}$  was branched off achieved an energy imbalance of  $-0.02 \text{ W m}^{-2}$  implying essentially perfect equilibrium. Continued integration of  $E^{280p}$  will lead to further reduction in the TOA energy imbalance.
3. The ocean model in *Gent et al.* [2011] CCSM4 PI simulation was configured with both (i) the overflow parametrization and (ii) tidal mixing scheme turned on. As we discuss in our manuscript, and which we further discuss below in response to the reviewer's concerns, these two options of the ocean model are tuned to the present day and as such should not be used for palaeo climate simulations [see *Peltier and Vettoretti* [2014], *Vettoretti and Peltier* [2013]]. But turning off these parameters in the Pliocene simulation, while keeping them active in the PI control would make the assessment of anomalies between the modern and the Pliocene simulation difficult due to this incompatibility between the two ocean models. It is for this reason that we have turned off both tidal mixing and overflow parametrization in our modern control, and therefore this is another difference between our PI control and *Gent et al.*.
4. There are some differences between the ocean mask in our PI control and that in *Gent et al.* [2011]. See *Vettoretti and Peltier* [2013] for details of the differences.

P 1, Line 9-10: This is a different way in which the CCSM4 has been used with major changes to the ocean vertical mixing physics, so not just boundary conditions are different.

The reviewers comment here is with regards to the statement in the abstract “With the new boundary conditions, the CCSM4 model simulates a mid-Pliocene which is more than twice as warm as that with the boundary conditions used for PliomIP Phase 1.” The reviewer believes that this statement does not include some of the other changes (such as the change from the ocean mixing profile) that also distinguish the results of our simulation from those of *Rosenbloom*

*et al.* [2013]. With regards to this concern, we present the table below which compares the temperature anomaly of the Pliocene simulation with POP1 style profile of diapycnal diffusivity, namely,  $\Delta TEoi^{400}_P$  with both controls (POP1 variants), as well as the anomaly of the 400 ppmv Pliocene simulation with the pelagic value of the diapycnal diffusivity ( $\Delta TEoi^{400}$ ) with both controls (in this case the controls also employ the pelagic value of the diapycnal diffusivity). This comparison shows that the temperature anomalies for both variants of the Pliocene simulation are very similar and that the effect of the POP1 mixing profile is to increase the anomalies by 0.3 degrees. Both of the anomalies with respect to PI control (3.8°C and 3.5°C) are roughly twice as large as the anomaly simulated by Rosenbloom *et al.*, (who obtained an anomaly of 1.8°C) with the CCSM4 model and PlioMIP1 boundary conditions. This analysis demonstrates that the claim that with the new boundary conditions “the CCSM4 model simulates a mid-Pliocene which is more than twice as warm as that with the boundary conditions used for PlioMIP Phase 1” holds true despite the nature of the mixing profile. This is an important result on its own and discussion of it is included in our revised manuscript.

	Anomaly w.r.t 280 ppmv control	Anomaly w.r.t 400 ppmv control
$\Delta TEoi^{400}_P$	3.8	1.8
$\Delta TEoi^{400}$	3.5	1.5

Since the reviewer has commented upon the impact of changes to the model, particularly in the ocean component, an additional point needs to be clarified. All our models have been run with the tidal mixing parametrization turned off. As per the discussion in our manuscript and in *Peltier and Vettoretti* [2014] this spatially heterogeneous field is tuned to the specific details of the present-day tidal regime and whose validity during time periods when the coastlines, bathymetric roughness, and eustatic sea level would have been different is questionable. Therefore, this parametrization has been turned off for the Pliocene simulations, and since the Pliocene simulations will need to be compared to the control simulations, in order to simplify this comparison without the subtlety of different mixing schemes, the parametrization has also been turned off in the controls. Our approach should be contrasted with that taken by *Rosenbloom et al.* [2013] with the CCSM4 model for PlioMIP. Although, the authors were careful to maintain consistency regarding the nature of mixing schemes between their control and Pliocene, they were nevertheless, forced to use the present day tidal mixing scheme within their Pliocene simulations because they opted to branch their Pliocene simulation from the existing PI control of *Gent et al.* [2011] which was configured with the present-day tidal mixing scheme.

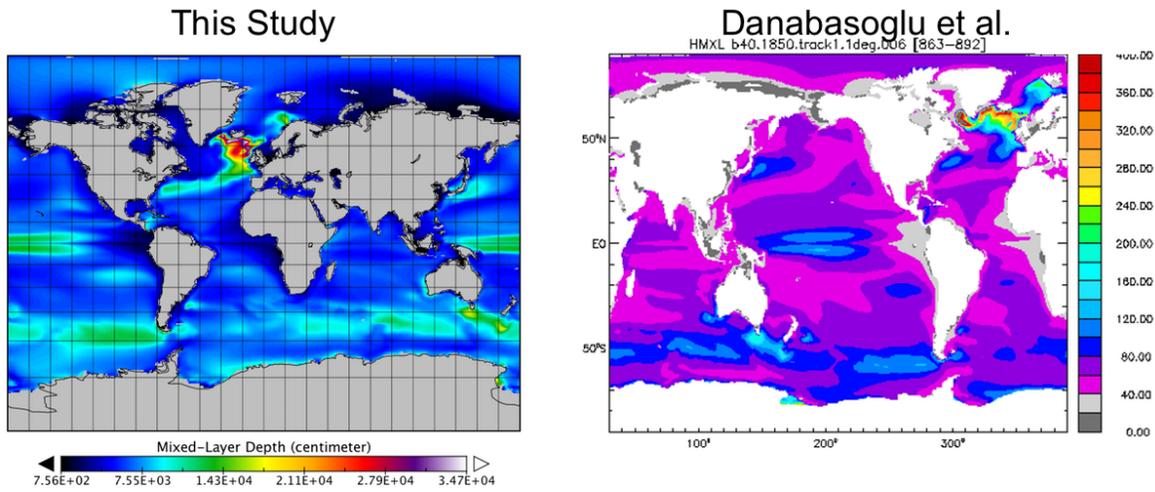
P 3, Lines 22-25: Cite Haywood et al (2011) when talking about the designs and aim of PlioMIP1

This reference will be included in the revised manuscript. Thank you for reminding us of this.

P 6, lines 10-15: Please discuss the choice of the ocean convective parameterization. Was the KPP scheme used apart from turning off the tidal mixing component? Was tidal mixing also turned off when the Kv was kept constant? Perhaps show your preindustrial POP1 simulation’s mixed layer depths. How is the 2-d distribution of mixed layer depth different

from the standard CCSM4 preindustrial simulation as described in Danabasoglu et al., *J. Climate*, 25, 2012?

The KPP convective parametrization was used in all simulations. The tidal mixing scheme was turned off in all simulations regardless of the choice of diapycnal diffusivity. The next figure compares our annual average mixed layer depth to that from *Danabasoglu et al.* [2012]. Our simulation is characterized by a robust mixed layer region in the north Atlantic and the GIN seas. The differences in the Labrador sea are expected based on the fact that overflow parametrization in our simulations was turned off (see *Danabasoglu et al.* [2012]).



P 7, line 8: The default 1° POP2 bathymetry field (KMT) is not from GTOPO30 as described here. The CCSM4 POP2 KMT field was generated from ETOPO2v2 as described and cited by Danabasoglu et al. 2012.

We thank the reviewer for bringing this to our attention. Comment to this effect has been included in the revised manuscript.

P 8, 18-20: This line, describing how the global mean salinity is adjusted for glacial climates to reflect a drop in sea level, and hence ocean volume, is not needed here as there is no glacial climate simulation. Thus, to make the manuscript more concise and focused, it should be deleted.

We believe that this statement, which clarifies the typical magnitude of ESL change for which a salinity adjustment becomes important, is pertinent to the discussion as to why a salinity adjustment was not performed and provides the necessary context for a reader who might be otherwise unfamiliar with the nature of such an adjustment.

P9, line 22: By what measure have the global mean SSTs come into equilibrium? Trends are still quite apparent in Fig. 3a.

The reviewer is, understandably, mistaken concerning an apparent SST trend. The reviewer is attempting to judge by eye the trend and there is a concavity in SST time series in the Pliocene simulations around year 1500. This is apparently giving the mistaken impression of a significant trend. However, numerical computation over the last two-hundred years of the simulation yields an SST trend of only  $0.02^{\circ}\text{C}/\text{Century}$  for  $Eoi^{400}_P$ ,  $0.04^{\circ}\text{C}/\text{Century}$  for  $Eoi^{450}_P$ ,  $0.01^{\circ}\text{C}/\text{Century}$  for  $E^{280}_P$  and  $0.01^{\circ}\text{C}/\text{Century}$  for  $E^{400}_P$  (also see table on the next page of this document). These values support our claim that the global mean SSTs have come into equilibrium, particularly for  $Eoi^{400}_P$  which is our “the Pliocene” simulation and for the control simulations. Because we now recognize that there is a possibility for a reader to mistakenly infer a larger trend than that which actually exists solely based on the figure, in the revised manuscript we now explicitly cite the numerically determined SST trends in the discussion of the equilibrium nature of our simulations.

P9, line 24: “...increases the rate of warming for the middle and lower and decreases the rate of warming for the upper ocean.” This decrease/increase of trend is not apparent apart from the very beginning of the start of the POP1-Kv runs. The trends at the end of the POP1-Kv runs look very similar to the trends at the end of the POP2-Kv runs. Typically there is a flattening as the ocean model equilibrates, but this is not the case here. Comparing Figs 3b-d, there does seem to be a transfer of heat from the upper ocean at depths above 550 m which leads to cooling at the beginning of the POP1-Kv runs, while the deep ocean below 1850 m starts to warm. There was probably not a net heat loss to the atmosphere at the start through venting the upper ocean, otherwise the global mean temperature (Fig. 4) would show a cooling trend at this time instead of a warming trend.

We thank the reviewer for bringing to our attention an error in the text of the manuscript. What line 24 on page 9 should have said is: “This increases the rate of warming of the lower ocean and decreases the rate of warming for the upper and middle ocean.” In the immediate centuries after the introduction of the POP1 profile of vertical mixing, there is vigorous vertical re-distribution of energy within the ocean. This leads to a cooling of the upper and the mid-depth oceans and warming of the abyssal ocean, which is expected on physical grounds. After this immediate response, the temperatures in the upper and mid-depth ocean continue to evolve with roughly the same trend as that before the introduction of the POP1 profile. In the lower ocean, however, the trends are characterized by a somewhat larger increase after the change.

P9, line 29: “...given the deep ocean sufficient time to come into equilibrium.” Again, Fig 3d shows that the deep ocean is far from equilibrium with major drifts by the end of the simulations. P9, lines 30-32: Trends of global mean temperatures are twice as large ( $0.07\text{K}/\text{century}$ , shown in Fig. 2) as what is shown in Rosenbloom et al. 2013 for the CCSM4 PlioMIP1 simulation, which is around  $0.03\text{K}/\text{century}$ . P13, lines 12-13: “...run to near statistical equilibrium...” By what measure are these simulations in a statistical equilibrium? As discussed above, deep ocean temperature trends are large and the global

mean ocean temperature at the end of the runs exceeds the CCSM4 PlioMIP1 simulation discussed in Rosenbloom et al. 2013.

We expect that the trends of the global ocean temperature for different vertical levels are small enough that the simulations can be considered equilibrated. The table below shows the temperature trends in units of  $^{\circ}\text{C}/\text{Century}$  for different parts of the ocean computed over the last two hundred years. Note that the global ocean trends presented here for all simulations have been updated from a trend of  $0.07^{\circ}\text{C}/\text{Century}$  for  $Eoi^{400}_p$ ,  $0.08^{\circ}\text{C}/\text{Century}$  for  $Eoi^{450}_p$ ,  $0.04^{\circ}\text{C}/\text{Century}$  for  $E^{280}_p$  and  $0.07^{\circ}\text{C}/\text{Century}$  for  $E^{400}_p$  that was reported in the original manuscript. The updated trends will be incorporated in the revised manuscript, and we have also incorporated the table below into our manuscript. In this table the regions upper, middle and lower ocean are defined as in Figure 3 of the manuscript.

	Global Ocean	SSt	Upper Ocean	Middle Ocean	Lower Ocean
$Eoi^{400}_p$	0.05	0.02	0.04	0.05	0.05
$Eoi^{450}_p$	0.05	0.04	0.03	0.05	0.06
$E^{280}_p$	0.03	0.01	0.01	0.02	0.05
$E^{400}_p$	0.06	0.01	0.03	0.05	0.07

Our work demonstrates the necessity of performing longer simulations with the revised PlioMIP2 boundary conditions than was necessary in the original PlioMIP, which did not include changes to the bathymetry and major oceanic gateways. Since no other group participating in PlioMIP2 has yet published results with the new boundary conditions fully implemented (*Kamae et al.* [2016] used the standard boundary conditions set which did not require changes to the land-sea mask) we suggest that our results, might be used by other groups to estimate how long they will need to run their own simulations to reach an adequate equilibrium.

The reviewer has correctly pointed out that the global mean temperature trend in *Rosenbloom et al.* [2013] Pliocene simulation was  $0.03^{\circ}\text{C}/\text{Century}$  while that in our Pliocene simulation the trend is  $0.07^{\circ}\text{C}/\text{Century}$  (older value from the manuscript, revised value in table above). At first glance, it may seem curious as to why the 500 year simulation of *Rosenbloom et al.* settled to a smaller ocean trend than our much longer simulation. However, there are several important reasons as to why their ocean model has reached equilibrium in such a short period of time:

1. The PRISM3 boundary conditions in PlioMIP did not specify any changes to the ocean bathymetry. This meant that the ocean in coupled-climate PlioMIP simulations that were initialized from present day temperature and salinity, or in models initialized from existing modern day runs (such as *Rosenbloom et al.*) were not characterized by the existence of a strong starting transient due to the modified bathymetry and circulation, or by the need to develop consistent T/S fields as required of the equilibrium state.
2. In addition to this, *Rosenbloom et al.* did not incorporate the PRISM3 changes to the land sea mask in the region of West Antarctica which is also expected to have aided in the faster approach to equilibrium.
3. Furthermore, the CCSM4 simulation of *Rosenbloom et al.* was branched at year 801 of the 1300 year PI control simulation. Given that their Pliocene simulation did not include

any changes to the bathymetry (which is partly because the PRISM3 dataset doesn't prescribe bathymetry, and partly because the authors did not implement the changes near Antarctica), save for the infilling of the shallow Hudson Bay region, this meant that each grid cell in the ocean was initialized with a robust and consistent temperature and salinity distribution, velocities, and circulation patterns. Including this 801 year PI simulation, their ocean model was actually integrated for 1300 years. In our case on the other hand, the changes that were necessary to apply the PlioMIP2 boundary conditions meant that several new ocean grid cells had to be created throughout the ocean, and therefore we've had to start our ocean from a state of rest (i.e. the velocities had to be completely spun up) and with interpolated temperature and salinity values for oceanic grid cells that exist on the Pliocene grid, but not on the modern day grid. This configuration requires longer integration to reach equilibrium.

Given the additional challenges that had to be overcome to get a fully PRISM4 compatible PlioMIP2 simulation going, it is remarkable that we have been able to get the trend in our Pliocene simulation down to  $\sim 0.05^{\circ}\text{C}/\text{Century}$ .

P13, lines 18-34: Please indicate over what region of the water column and latitude range the maximum Atlantic Meridional Overturning Circulation metric was calculated. One needs to exclude the upper 500m or so of the wind-driven tropical cells.

The MOC maximum has indeed been computed while excluding the wind-driven tropical cell.

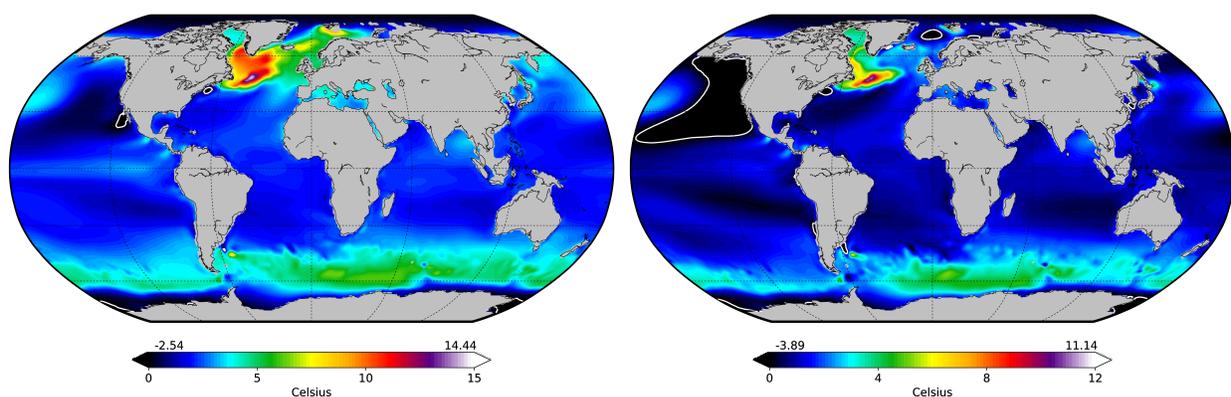
Figure 5 and Table 2 temperature both say Mean Annual Surface Air Temperature, but in comparing the same runs, the values do not agree. Table 2 numbers are higher than shown in Figure 5, suggesting that Table 2 numbers are TS, the Surface Temperature (radiative) which tends to be higher than the TREFHT or 2m air temperature typically used for 'MASAT'.

We thank the reviewer for pointing out this discrepancy. The values in Tables 2 and 3 are derived from the TREFHT field of the atmospheric component. These values are those that should be used for intercomparison with other studies. Due to technical issues that arose during the post processing of the original monthly history files that the model writes, the TREFHT variable did not get processed into time series for most models for the first several hundred years. Therefore, the bottom most hybrid level of the atmospheric temperature field was used to create the time series shown in manuscript Figure 5. The difference between the globally and yearly averaged values of these two variables only differs by a few tenths of a degree, while there is no change in the evolution of the time series. It is for this reason that Figure 5 shows the evolution of the globally and annually averaged temperature of the atmospheric layer closest to the ground. We ought to have made this clear but unfortunately this caveat didn't make it into the first draft. It has been made clear in the revised manuscript.

Figure 10: It is difficult to discern the contours in 10b-c. There appears to be possibly a very large cold bias in c and a smaller one in b, based on the minima shown in the color bars for these two panels of anomalies. There appears to be a large black swath in panel

c emanating from the eastern boundary in the N. Pacific, suggesting a very large cold anomaly with respect to the Modern control.

The SST anomalies in manuscript Figure 10b and 10c show the Pliocene - control anomalies with respect to each of our controls. There is a region along the eastern boundary of the N. Pacific in the anomaly field with respect to 400 ppmv control over which the 400 ppmv Pliocene is cooler than 400 ppmv control. This was meant to have been clearer in the original figures, but the zero-anomaly contour line in our code was set to black which did not show up against the color scheme. We have clarified this with updated figures which show a zero-anomaly contour in white as shown below.



## References

- Bitz, C.M., Shell, K.M., Gent, P.R., Bailey, D.A., Danabasoglu, G., Armour, K.C., Holland, M.M., Kiehl, J.T., 2012. Climate Sensitivity of the Community Climate System Model, Version 4. *J. Climate* 25, 3053–3070. doi:10.1175/JCLI-D-11-00290.1
- Cohen, J., Jones, J., Furtado, J. and Tziperman, E.: Warm Arctic, Cold Continents: A Common Pattern Related to Arctic Sea Ice Melt, Snow Advance, and Extreme Winter Weather, *oceanog*, 26(4), 150–160, doi:10.5670/oceanog.2013.70, 2013.
- Cunningham, S. A., Kanzow, T., Rayner, D., Baringer, M. O., Johns, W. E., Marotzke, J., Longworth, H. R., Grant, E. M., Hirschi, J. J. M., Beal, L. M., Meinen, C. S. and Bryden, H. L.: Temporal Variability of the Atlantic Meridional Overturning Circulation at 26.5°N, *Science*, 317(5840), 935–938, doi:10.1126/science.1141304, 2007.
- Danabasoglu, G., Bates, S.C., Briegleb, B.P., Jayne, S.R., Jochum, M., Large, W.G., Peacock, S., Yeager, S.G., 2012. The CCSM4 Ocean Component. *J. Climate* 25, 1361–1389. doi:10.1175/JCLI-D-11-00091.1
- Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model Version 4, *J. Climate*, 24, 4973–4991, doi:10.1175/2011JCLI4083.1, 2011.

- Haywood, A. M., Hill, D. J., Dolan, A. M., Otto-Bliesner, B. L., Bragg, F. J., Chan, W.-L., Chandler, M. A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D. J., Abe-Ouchi, A., Pickering, S. J., Ramstein, G., Rosenbloom, N. A., Salzmann, U., Sohl, L. E., Stepanek, C., Ueda, H., Yan, Q. and Zhang, Z.: Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project, *Clim. Past*, 9(1), 191–209, doi:10.5194/cp-9-191-2013, 2013.
- Howell, F.W., Haywood, A.M., Otto-Bliesner, B.L., Bragg, F., Chan, W.-L., Chandler, M.A., Contoux, C., Kamae, Y., Abe-Ouchi, A., Rosenbloom, N.A., Stepanek, C., Zhang, Z., 2016. Arctic sea ice simulation in the PlioMIP ensemble. *Clim. Past* 12, 749–767. doi:10.5194/cp-12-749-2016
- Kamae, Y., Yoshida, K. and Ueda, H.: Sensitivity of Pliocene climate simulations in MRI-CGCM2.3 to respective boundary conditions, *Clim. Past*, 12, 1619–1634, doi:10.5194/cp-12-1619-2016, 2016.
- Levitus, S. and Boyer, T. P.: World Ocean Atlas, Volume 4: Temperature NOAA Atlas NESDIS 4, U.S. Government Printing Office, 1994.
- McCarthy, G., Smeed, D., Johns, W., Frajka-Williams, E., Moat, B., Rayner, D., Baringer, M., Meinen, C., Collins, J., and Bryden, H.: Measuring the Atlantic meridional overturning circulation at 26°N, *Prog. Oceanogr.*, 130, 91–111, 2015
- Overland, J. E., Wood, K. R. and Wang, M.: Warm Arctic-cold continents: climate impacts of the newly open Arctic Sea, *Polar Research*, 30(0), 4045–14, doi:10.3402/polar.v30i0.15787, 2011.
- Peltier, W. R. and Vettoretti, G.: Dansgaard-Oeschger oscillations predicted in a comprehensive model of glacial climate: A “kicked” salt oscillator in the Atlantic, *Geophys. Res. Lett.*, 41(20), 7306–7313, doi:10.1002/2014GL061413, 2014.
- Rosenbloom, N. A., Otto-Bliesner, B. L., Brady, E. C. and Lawrence, P. J.: Simulating the mid-Pliocene Warm Period with the CCSM4 model, *Geosci. Model Dev.*, 6(2), 549–561, doi:10.5194/gmd-6-549-2013, 2013.
- Salzmann, U., Haywood, A. M., Lunt, D. J., Valdes, P. J. and Hill, D. J.: A new global biome reconstruction and data-model comparison for the Middle Pliocene, *Global Ecology and Biogeography*, 17(3), 432–447, doi:10.1111/j.1466-8238.2008.00381.x, 2008.
- Sun, L., Perlwitz, J. and Hoerling, M.: What caused the recent “Warm Arctic, Cold Continents” trend pattern in winter temperatures? *Geophys. Res. Lett.*, 43(10), 5345–5352, doi:10.1002/2016GL069024, 2016.
- Vettoretti, G. and Peltier, W. R.: Last Glacial Maximum ice sheet impacts on North Atlantic climate variability: The importance of the sea ice lid, *Geophys. Res. Lett.*, 40(24), 6378–6383, doi:10.1002/2013GL058486, 2013.

Zhang, Z., Nisancioglu, K.H., Chandler, M.A., Haywood, A.M., Otto-Bliesner, B.L., Ramstein, G., Stepanek, C., Abe-Ouchi, A., Chan, W.-L., Bragg, F.J., Contoux, C., Dolan, A.M., Hill, D.J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Rosenbloom, N.A., Sohl, L.E., Ueda, H., 2013. Mid-pliocene Atlantic Meridional Overturning Circulation not unlike modern. *Clim. Past* 9, 1495–1504. doi:10.5194/cp-9-1495-2013