I have studied the reviewers comments as well as the author response carefully.

The paper presents a great deal of hard work and represents a very significant scientific contribution to PlioMIP2. It is clear that the paper will be worthy of publication in the PlioMIP2 CP special issue subject to the satisfactory completion of specific modifications/clarifications and additions. Whilst I totally understand the desire to robustly defend your own work, I would like to respectfully add that it is the authors responsibility to provide analyses and figures that satisfy the reviewers. This seems to have been a particularly important issue for this paper given the different results presented compared to the published and standard version of CCSM4 (by Rosenbloom et al. etc). I would also like to see the wording of the paper (in the title) altered to make it clear within the PlioMIP2 project that these results are derived from an altered version of CCSM4 (compared to the standard NCAR version). Perhaps something like “Regional and global climate for the mid-Pliocene using the University of Toronto version of CCSM4 and PlioMIP2 boundary conditions” would suffice.

Specific comments/requests.

1. Pre-industrial control

Given the reviewers comments the pre-industrial control should be validated against observations, where available. For example, SST and sea ice versus HadiSST. The authors can refer to the Climate Data Guide to find relevant observational datasets (https://climatedataguide.ucar.edu/).

A standard approach in climate modelling is that first a model simulation (PI or PD) should be validated against observations for the regions or variables being studied before analysing sensitivity simulations, particularly in this study where it seems parameterizations have been modified, or turned off, in the PI control. I appreciate and understand the authors’ comments and views on this matter but would also point out that the reasoning will not be intuitive to many that one should change or turn-off modern-based parameterizations as not appropriate for the Pliocene, but then also change or turn-off for the preindustrial where they are appropriate and well-tested in terms of enabling the model to perform well against observations. This discussion and relevant figures can be added as supplementary material of the paper and thus not disturbing the overall flow of the main paper.

2. Pacific Warm Pool

More justification for the expansion of the warm pool in the Eoi400 simulation seems to be required. Given the reviewer comments it appears to be important to distinguish whether the expanded warm pool is due to one or more of the new PlioMIP2 boundary conditions or rather the changes to the ocean parameterizations. Since the authors have been commendably comprehensive in performing the factorization simulations of PlioMIP2, they can provide a more detailed analysis to let the reader understand which of the PlioMIP2 forcings/boundary conditions (vegetation, different paleogeography, CO_2) and/or the different POP parameters lead to this result. Note also that it is a little difficult to see an expansion of the warm pool in Fig. 10 from the panels given. Figure 10 could be redrafted to include contours every ~1°C, including negative anomalies in b) and c) and centred on the Pacific. Also, the large area of cooling in Eoi400 relative to the E400, which is somewhat obscured in Fig. 10c by showing only positive anomalies, needs some form of explanation.

3. Sea ice

On page 4, of the authors' response, it is suggested that the sea ice climatology for the PI control period used in the Rosenbloom et al. 2013 PlioMIP1 comparison may have been different than for the period described in Gent et al. 2011. However, as reported in the atmospheric diagnostics on the publicly available web site: http://www.cesm.ucar.edu/experiments/cesm1.0/, the mean annual NH sea ice area is 11.73 x 10^6 km^2 for the period documented in Gent et al. 2011, which is approximately the same as shown in Table 2 of Rosenbloom et al. 2013.

I suggest comparing the sea ice predictions for the PI to the HadiSST preindustrial period. If the sea ice extent is greater in this region compared to HadiSST this could, when compared to a Pliocene simulation without sea ice in that region, yield a large warm anomaly, which is just a function of a different sea ice result in the PI (compared to other PI CCSM4 simulations).

On page 5 of the authors' response, the authors may be confusing the sea ice extent reported in Howell et al. 2016 with the sea ice area reported in Rosenbloom et al. 2013 for the CCSM4 PI. They show the annual cycle of sea ice area of their PI simulation (left), vs. Howell et al. 2016 (fig. 2) sea ice extent in the PlioMIP1 group (right). Sea ice extent will be larger than sea ice area. Sea Ice extent is computed by integrating over all grid boxes with sea ice, weighting by total area of the grid box, regardless of the fraction of ice within the grid box, thus obtaining an estimate of the size or extent of the ice pack. On the other hand sea ice area is computed by integrating area after weighting each grid box by the fraction covered by sea ice. Thus, the PI control mean monthly sea ice area climatology does not appear to be comparable to the sea ice extents reported in Howell et al. 2016. This should be re-examined and corrected as necessary.

4 North Atlantic SSTs

Otto-Bliesner et al., GRL, 2017 have shown that closing of the Arctic gateways improved the simulation of North Atlantic SSTs between 40-60N as compared to the earlier CCSM4 PlioMIP1 simulation. Please cite this paper.

If the pre-industrial simulation (E280) in this paper does have more sea ice in the Labrador Sea in summer than might be indicated by HadiSST, this needs to be clearly stated when discussing warming in the North Atlantic, E400 - E280, Eoi400 - E280.

Figure 11: Brierley et al. proxy-inferred SSTs are for early Pliocene. In fact in their Table S1, they include a temporal correction for those data that have the averaging interval centred in the mid-Pliocene in order to create their early Pliocene estimates. The PRISM3 data compilation needs to be used in this figure and only data points in the Brierley compilation that are in the PRISM3 time interval included since your paper is on the mid-Pliocene/PlioMIP2 paper.

5 Winter temperature response over NH continents

The widespread cooling over the Northern Hemisphere continents during DJF is not supported by the proxy data. As such, more analysis of this feature needs to be included in the paper to understand such a response.

Potential figures that could be referred to from the main text to the suppl section include maps of TLAI, snow cover, and circulation changes with the same NH projections as Figure 7. A comparison to previous model results from PlioMIP1 may also help. The PlioMIP1 models with prescribed vegetation use the same vegetation reconstruction and similar Greenland ice sheet as PlioMIP2. At least some of the PlioMIP1 models show much reduced Arctic sea ice, which is relevant to the authors' hypothesis that the winter cooling is related to a reduced Arctic sea ice in the presented simulation. The authors can document which, if any, of the PlioMIP1 models also show widespread winter cooling over the NH continents.

6. Equilibrium Climate Sensitivity (ECS) and Earth System Sensitivity (ESS)

It is very important to define what kind of sensitivity being examined very carefully. Table 2 in the paper presents the ECS (last column) of 7.4 and 6.3K. The manuscript also uses the term 'ECS' to discuss Table 2 on page 10. Is this instead ESS? If truly ECS, then these values are much greater than in the standard version of CCSM4. Comparing E400 to E280, one can estimate an ECS of ~4K, which is larger than the standard version of CCSM4 suggesting that the differences made to the ocean physics may have increased the ECS of this version of CCSM4. Perhaps it has also increased the ESS. Please clarify in the revised paper.

7 Figure legends

Please define details of averaging regions etc.

8 Spin up of the PI control

When asked for details about the pre-existing PI spin up of 3500 years length the authors reference Vettoretti and Peltier, 2013 (hereafter VP) on page 7 of the authors' response. However, VP describes a 3200 year long preindustrial spin up using CCSM3, not CCSM4. This detail of a change in model has not been included in the authors' response nor the manuscript. However, Fig. 2 in VP suggests the preindustrial spin up in VP may not be the one used here. Fig. 2c in VP, shows the evolution of sea ice area in the NH which is very smooth ending at year 3200 at ~11 x 10^6 km^2. What is shown on pg. 7 of the authors' response, in contrast, is an evolution of NH sea ice 'extent' (is it 'extent' or area?) that shows an abrupt increase at about year 750 rising from ~11.7 x 10^6 km^2 to ~14 x 10^6 km^2, hence not looking like the smooth evolution in VP, fig. 2c. What happened at year 750 to cause the relatively abrupt increase in sea ice “extent” or area? Is the jump coincident with a growth of sea ice in the Labrador Sea? This then seems to persist into the later period of comparison. Another manuscript, Peltier and Vettoretti, GRL, 2014, hereafter PV, rather, uses CCSM4 for a long preindustrial control of length ~2863 years, still not 3500. Of the 2863 years, according to the PV paper, the last 1200 years had the tidal mixing and overflows turned off, POP1 Kv etc. like as described for the simulations occurring after the forks in Fig. 3-5. Did the authors go back and forth on the overflows and mixing schemes in the PI control as this suggests? Please give details of the spin up, and validation of the PI simulation over the period of comparison, either in the main text or Supplementary Information.