

Author's response to Referee #2

Referee Comment (RC): The paper of Baatsen et al. presents results from two simulations done using a recent version of the NCAR model (CMIP-5 like ?) and the late Eocene paleogeography they developed in an older paper. While I think these simulations worth to be published, an important work of re-writing is required before making this manuscript easy to follow and fluid.

This manuscript suffers from two issues; 1) many vague statements, story telling not always supported by the figures are found and 2) too many subjects are touched upon. I suppose that in their willingness to be sufficiently exhaustive, the authors have forgotten to keep focused their paper. I tried below to follow the paper and to suggest some corrections and/or some parts to be deleted.

I suggest to the authors to divide their manuscript into two papers. In the first one, they could leave all parts on the general description of the results and the model-data comparison. In the second one, they can focus on the climatic sensitivity and the super-rotation. Following that recommendation, they may add more info to this part which is not always very convincing. It is one possibility; they may find other means to arrange the manuscript.

Author's Response (AR): *We would like to thank the reviewer for the extensive and in depth comments of the manuscript. We agree to shorten the paper mainly in terms of figures/tables and cutting back on the description of model. The structure and language of the manuscript will both be improved to increase readability and clarity.*

RC: In addition, while I do appreciate the work achieved by Baatsen et al. in reconstructing the late Eocene geography, they have to be more balanced and to recognize that many uncertainties exist and that other paleogeographies may be tested (see all papers by Ran Zhang, Zhongshi Zhang and its colleagues). Here, it looks like you have the ultimate paleogeography and all others are worst than yours; this is not the case and you should be more cautious. In addition, even if one accepts your paleogeography as being the most accurate; another uncertainty is the paleotopography. For example, topography over the Himalayan - Tibetan plateau area is badly constrained and may strongly impact both global climate and ocean dynamics (see paper in *Climate of the Past* by Su, B., Jiang, D., Zhang, R., Sepulchre, P., and Ramstein, G).

AR: *The aim here is to show that with increased model resolution and the most recent information on palaeogeography, a better representation of the middle-to-late Eocene climate can be achieved. It is certainly not the intention to claim that the model set-up and reconstruction are perfect, so this will be better nuanced. Since the geography reconstruction was made based on a Paleomag rather than Hot spot plate-tectonic reference, it also serves to show the array of possible uncertainty regarding geographical boundary conditions rather than claiming that one is better.*

RC: To conclude on the climate sensitivity. The model simulates a global warmth of 10°C from the CTRL run to the 2xPIC. Many factors are responsible for it; land-ice, vegetation cover and paleogeography. Without performing factorisation as already done by Lunt and colleagues, it remains hard to attribute a clear reason to what is occurring here. Conversely,

the model simulate a 3°C warming between 4xPIC and 2xPIC, here the climate sensitivity is the classical one of the NCAR family model. So, I am left puzzled with the references to higher spatial resolution and to change in atmospheric grid to explain your results.

AR: *Both the resolution and atmospheric grid used here were previously shown to increase the sensitivity of the CESM model. Clearly the model is also experiencing a significant heating from the Eocene boundary conditions, which is more substantial than in most previous studies but similar to the ~5 C of Caballero and Huber (2013). This will be discussed in more detail and additional analysis regarding the sensitivity and temperature offset will be carried out.*

RC: Specific/major Comments: l. 88 not entirely true, see Licht et al. (Nature, 2014). I do understand the need for clarity when writing a paper. In that sense, Baatsen et al. simplify the story and claim that no Late Eocene paleogeography have been used before them. But this is not true, Licht et al., in 2014, have also tested a Late Eocene paleogeography, not for the same goal but still . . . please modify the paper accordingly.

AR: *Thanks for pointing out this shortcoming. A 40Ma reconstruction was indeed used by Licht et al. but there is little information on how it was created. They also use a coarse resolution coupled model (FOAM) to drive an atmosphere-only simulation, which is not really comparable to the simulations presented here.*

RC: l. 94 I dont remember accurately conclusions reached by Beerling et al. in their 2011 PNAS paper but should be nice to state here the degree of global warming potentially reached when doubling and/or quadrupling methane concentration during the Eocene as it is what they did in their 2011 paper.

AR: *According to Etminan et al. (2016), the contribution of CH₄ is limited and makes the radiative forcing equivalent to that of 2.15x and 4.69x PI CO₂, respectively, as mentioned later in the paper. This will be mentioned now earlier in the paper for clarity.*

RC: l. 110 - 90 x 144 correspond to 2° in latitude and 2.5° in longitude.

AR: *This was a typo which will be corrected; the model grid should be 96x144 which does correspond to 1.9x2.5deg*

RC: l. 111 - avoid vague statement such as 'generally tropical ... and mixed forests further poleward'. Are you imposing zonal band of vegetation? in addition, are you using phenological models of vegetation, i.e. interactive Leaf Area Index? or is is an only PFT models by which the climate is influenced by biophysical parameters such as albedo, roughness and evapo-transpiration?

AR: *The implementation of vegetation is rather simple and consists of static PFT's determining their response in the model. This should be treated in more detail within the model set-up and maybe demands a figure showing model geography and vegetation.*

RC: Figure 1 - While the 2x simulation is close to equilibrium, the 4x would require more integration time before reaching the equilibrium as testified by Fig. 1b. So, lines 140-144 are misleading. One can also think that the reduced vertical mixing in the 4xPIC case is due to the still non equilibrated deep ocean. In addition, Figure 2b shows that age tracers are more equilibrated for the 4xPIC case than for the 2xPIC case. This is inverse to what is shown by Figure 1b. Why that ? Finally, the constant trend of THC index (l. 148) is hard to be interpreted because ocean dynamics can behave non linearly and can show tipping point after several hundred years of constancy. I like the way the authors do carefully treat the question of the equilibrium in their two simulations. It is something which matters a lot. However, putting more and more diagnostics can be also troubling ! So, my last comments are more questions open to every one in the modeling community than a direct criticism of the author's work.

AR: *The 4x PIC is indeed showing larger trends than the 2x PIC one, but the scaling in the figure make those appear quite dramatic. Looking at the values in Table 2, there is about 0.5C warming per 1000 years which is quite small and in fact similar or less than what is seen in most other studies (e.g. 0.016C/century in Goldner et al. 2014). It should be acknowledged that the model is not yet fully equilibrated here, but to get rid of this trend would require several 1000s more years of model integration. The time series of MOC strength show that there is no trend over the last 1000 model years, so there seems to be no link to slowly warming deep ocean temperatures. The different trends in age tracers is caused by the deep mixing of younger waters opposed to ageing of stagnant regions; this will be mentioned in the revised paper.*

RC: Figure 3, this is an author choice but personally, I find the color palet very awkward and hard to look at. 3.1 If you describe your results, please dont mix with discussion?

AR: *The colour scales were designed specifically to be easy to interpret and friendly to people with mild colour blindness (they were also used in Baatsen et al. 2018). Maybe an adjustment of the scaling and density of contour lines can help improve the readability. This section will be restructured to have only a short presentation of general results and a discussion later on.*

RC: l. 170-175 you are referring to stratification but I do not see stratification indice on Figure 3 unless you are using the poorly visible contouring of the mixed layer depth? This discussion is not supported by your model outputs and following lines 175-177 are even worst. I suggest removing all these lines from 170-177 and move them to a discussion part.

AR: *Low stratification was linked to high upper level densities; this will be clarified. The discussion will also be moved further down.*

RC: l.181 - path and depth (200-500 m) > are you meaning that the depth of the DP in your paleogeography is from 200 to 500 m ? Not clear

AR: *This is regarding the model geography, adding a figure on this should help.*

RC: l. 185 where the reader should look at ?? to see the southward location of the polar front ??

AR: *The front is best seen in the barotropic stream function; this will be mentioned.*

RC: l. 191-193 very vague statement that your results compare well with those of Hutchinson based on what?

AR: *A figure will be added with a comparison to proxies and other model studies (with a description of the sources of the data).*

RC: l. 178 - 193 - modeling results from this study + data + modeling from others are mixed here . . . once again, this is a mess to follow.

Figure 3c, thick white lines show 0Pa – OK - but you also have a thick white line owing to the transition from negative to positive barotropic stream function. This is a bit harsh to follow. > l. 202 to l. 207 so what ?? can you discuss these results with those from organic data, is the amplitude the same ? etc . . .

AR: *Most of this part will be moved to the discussion section; the figure will be improved. The motivation to use 2x and 4x PIC forcing is to cover middle-to-late Eocene temperatures so it is relevant to mention a similar range is seen in estimates from benthic $d^{18}O$.*

RC: l. 196-209 - Back to global results in term of SST and still with a mix of data and model.

l. 210-240 - Global overturning functions and associated OHT

l. 215-217 - So What ??

AR: *This part will be re-written.*

RC: l. 221-222 - this sentence is not true at all. Figure 4d shows that more than a half of the OHT between the equator and 50°S (so not low latitudes but low to middle latitudes) are not due to the Pacific (compare red and black lines).

AR: *This will be adjusted to show Indo-Pacific rather than Pacific-only similar to the Pre-industrial case. The description will be re-phrased accordingly.*

RC: l. 226-230 - once again, these statements are not supported by your model results. From 50°S to south pole, all lines are converging. So, you can not write that the deep meridional overturning cell explains this feature in your paleo run as the OHT is the same in the modern while the overturning stream function is not !

l. 229-231 - So What ?

AR: *The difference related to Ekman upwelling in the ACC is quite clear, but the description should be modified as there are indeed no changes beyond ~55S. The main point remains that a significant redistribution of heat fluxes takes place from the NH towards the SH.*

RC: I. 243-259 Back to global feature and time evolution over the simulation Why here ??

I. 251-253 - Provide some details or just remove because it is a bit harsh written like that with no references.

AR: *These parts will be re-arranged.*

RC: I. 262-305 Data-Model comparisons - Can you provide an overall scoring using sum of square residuals for example ? it will be better than subjective term like mixed, substantial or good agreement. Another reason explaining data model disagreement may also be linked to the time period you are using. In 4 Myrs, CO₂ and orbital parameters may have changed, hence explaining data spreading not represented by your modeling.

AR: *The scarcity and quality of the available proxies do not allow the use of objective statistical analyses or shorter time intervals here. Proxies can be strongly biased or a slight shift in strongly regionally-dependent conditions can easily mess up such a comparison. The most transparent way is to just show the proxies with model results and leave to the reader the interpretation of individual points.*

RC: I. 271 Can you provide a reference for the saturation of UK37 above 28°C ?

AR: *Yes, e.g. Conte (2006), will be added.*

RC: I. 278-282 Remove these sentences, they bring nothing to the debate. As you are using a 4 per 4 ° box to generate temperature comparable to data; it is now hard to write that you have slight discrepancy in the position and/or model's limitation to resolve sharp upwelling zone (or at least for this point, you need to refer to previous studies showing that ...). Once again, avoid vague statement and/or non supported statement. The next sentence is not understandable (the existence of colder temperatures in the . . .).

AR: *This part will be rephrased. The model spatial variation overlaps with that of the proxies, so it can actually motivate the presence of nearby cooler waters.*

RC: I. 305-330 Figure 3 but with winter/summer features. Suggest to remove all this part.

1) plotting the zonally average summer TOA incoming solar radiation, you will see that there is a decrease around 60° and then an increase again till the south Pole. So, no need to solve any mysteries concerning the minimum summer temperatures in the center of polar gyres (if they are centered around 65°).

2) all the discussion referring to previous work by Bijl and Huber is hard to follow without a sketch pointing out where are the currents, the authors are referring to.

3) nothing new in writing that winter hemisphere has a deeper thermal gradient and thus winds and gyres.

AR: *As suggested, this figure will be removed along with most of the discussion.*

RC: I. 332-371 Comparisons with some previous runs. Suggest to remove all this part. Once again, many crucial numbers and references are lacking in this part making it vague and not very convincing. Can you state reference's papers for each version of the NCAR model you

are referring to ? you first write that the spatial resolution is not the same, and then, the initial conditions and then the numerical grid in the atmosphere and also the aerosols. What about the duration of each run ? And after all, why doing this comparison if there are so many things that have evolved since the GH14 paper . . . Figure 8 may be interpreted as showing two very different states of the climate and so what's the point here ? two very different version of the same model shows two very different climatic state ? Figure A3 also shows that the old version was simulating a latitudinal thermal gradient more flat than the new version (i.e. compare dashed red line and green line).

AR: *This figure will be removed and replaced by a more general overview of model results and proxies. The aim of this figure was to show the improvement being made in the general output and the added detail in the results, but it does not add much in general. There is only a slightly reduced thermal gradient in the Southern Hemisphere, which is the result of a weaker seasonal cycle (Figure A4) and the use of a completely closed Tasmanian Gateway in GH14.*

RC: I. 340-342 Can you provide a reference for such a statement ?

AR: *This is an observation from running the model in different configurations; there are no publications that make a 1-on-1 comparison under similar conditions.*

RC: I. 342-344 Conversely, comparison with results from Hutchinson is light whereas the paleogeography is the same. For example, you write 'similar global temperatures are found ...'(I. 343- 344). Can you give numbers mean global SST and air temperature for both models, yours and the one of Hutchinson ?

I.346-359 This is not the same geography and the CO₂'s sensitivity is not the same. So, no surprises, results are different !

AR: *As suggested above, this discussion will be removed and a better comparison will be given with the results of Hutchinson et al. (2018) in a new figure.*

RC: I. 374-400 General Description of the atmosphere.

I. 375-376 - Where are results from 2xPIC in this part ?

AR: *As mentioned, they are qualitatively similar and therefore only regarded in zonal mean plots and tables.*

RC: I. 403-426 General Description of the zonally averaged distribution of 2 m air temperature with 2x and 4x. Here many classical results are found by the authors . . . in particular about the polar amplification phenomenon. I am not sure that it brings anything new to the knowledge of the paleoclimate community. Could be remove to gain place and clarity.

AR: *The figure and description of zonal mean air temperatures will be removed and solely discussed in the comparison to proxies.*

RC: I. 429-467 Data model comparison over land

I. 471-472 So, you are not able to resolve data model disagreement as many authors suggest winter positive temperature in the continental interior of Siberia (see work of Upchurch, Valdes, Suan etc . . .)

I. 476-478 are you using interactive phenology in your set-up ? in this case, you should be able to show the LAI or the vegetation cover ? if not, your explanation is not possible. If yes, please show the simulated distribution vegetation cover (not the imposed one). If you impose a forest in an arid area, LAI should be 0 and the vegetation model should be able to predict no forests and you should have only barren soil.

AR: *The PFTs are prescribed, but how they respond is dependent on the climatic conditions. Most of the papers on the interior of Siberia seem to consider Early Eocene or late Cretaceous? There will be some more discussion on this issue in the revised paper.*

RC: I. 495-525 Zonal winds and super rotation . . .

AR: *The figure and description of the atmospheric zonal mean vertical cross sections will be taken out.*

RC: I.532 can you provide main principles of the Etminan method ? as the radiative forcings is a logarithmic function of greenhouse gas concentration, it is generally admitted that each CO₂ doubling will have the same influence (if not accounting for internal feed- backs), here you have 4.23 and 4.48, is the difference due to internal feedbacks of the Earth System Model?

AR: *The main reason for the nonlinearity is the interaction between different greenhouse gases; also the pure logarithmic behaviour is only the first-order contribution.*

RC: I.540-544 not sure to get your point here. If you are in a virtually ice-free world, how can, you have albedo effects due to snow (add seasonal polar view of snow and sea ice cover) ? is it more related to clouds feedbacks for example ? Again, concerning the growing season, this should be demonstrated and not just written. Or at least, expand your discussion here. Otherwise, it looks again as vague statement.

AR: *There is very little snow/ice, especially at 4x PIC but there is still winter snow cover. At 2x PIC there is still limited snow/ice cover but there is some influence on the climatic response that will be mentioned. The effect of vegetation and its response will be looked at in more detail.*

RC: I. 545-547 is it due to larger snow and sea-ice cover in our modern climate ?

AR: *Mainly the difference in snow/ice is thought to be the cause, but this is not discussed as there is no 2x PIC pre-industrial control experiment.*

RC: I. 558- 565 Interpreting OLR is always ambiguous as the thermal response does influence it. Here, you write that you will explain why you have a polar amplification but you end up by writing that you have more OLR because your high latitudes are warmer. No explanations of why they are warmer ? Or I have missed something ?

AR: *The OLR (clearsky) mainly shows the response to higher temperatures, but the main contribution is from clouds. This will be described more clearly.*

RC: In addition, I keep in mind from the Caballero and Huber PNAS paper that the super rotation strongly decrease the low cloud cover, hence leaving more solar radiation penetrating the atmosphere and warming it. Do you see that between 4X and 2X ?

AR: *As shown in figure 14, there are some reorganisations but no clear reduction in cloud cover between 2x and 4x PIC so this does not seem to be the case here.*

RC: Fig.14 f : Cloud cover, changes between -5 and + 5 %, looks very small and the same range than above (14d) ? Otherwise, it confirms what I suspect, i.e. more high clouds in your mid-to-high latitudes that may induce warmer high latitudes ? This is however possibly offset by changes in low cloud cover . . .

AR: *The changes in zonal mean cloud cover are small but significant as shown by the TOM fluxes. Indeed, increased high cloud cover is keeping middle/high latitudes warm as is reduced low cloud cover over the Arctic in summer.*

RC: I. 586-600 - not clear, please provide more details. Your first CO2 doubling (I. 590) has a larger effect because you also do account for changes in vegetation cover, land-ice, paleogeography etc . . . Then, you write that we can not add the radiative forcing but the, you do this addition assuming a linear behavior . . . How did you get ESS = 5.68- 8.97 ?? and 6°C per CO2 doubling ? (you have 3 °C in the Table 3 from 2xPIC to 4xPIC, or you have 10 °C for the first round but accounting for all geological factors)

AR: *The discussion regarding ESS/ECS and geography influences will be re-written accordingly.*

RC: I.635-652 - Please provide more details or at least, explain where you want to go with all these numbers.

AR: *Mainly this shows that the general results regarding S/G do not change using various methods, but it will be further clarified.*

RC: Summary and conclusions:

I. 685-690 Once again, these statements have not been demonstrated in this paper. Higher resolution is not the solution to everything and in case, you have to explain which physical processes are impacted by a higher resolution and how they explain your warmer climate simulation. The statement on the atmospheric grid is also doubtful - this is maybe true but please, provide explanations, what changes are occurring when modifying the grid. Why your more specific geography reconstruction should allow for a better representation of the Antarctic summer warmth and zonal heterogeneity in the southern ocean.

AR: *The higher resolution and new geography reconstruction mainly allows for a more regional interpretation of proxies to cause improved agreement on one hand and explain some discrepancies that remain. The effects of the grid and dynamical core will be discussed in more detail and further clarified.*

RC: To conclude, I am still left with many unanswered question, the main one is why you have this 10°C warming between your control run and between you 2xPIC ? it is probably due to a mixed of land-ice, vegetation changes and paleogeography effect. However, to provide a firm quantification of the effect of these boundary conditions, you have to do the factorisation method as done by Dan and others in the Miocene I guess. Conversely, your CO₂ model sensitivity remain in the lower range with 3°C (from 2xPIC to 4xPIC). Hence, I do not understand why you want to incriminate higher resolution and/or changes in atmospheric grids to explain your global warmth. You first have a large change in temperature when comparing CTRL to 2xPIC but then when comparing 4xPIC to 2xPIC, you have a more classical answer.

AR: *As stated, the model in its configuration used here responds strongly to the applied changes in boundary conditions but less so to increased radiative forcing from greenhouse gases. This will be looked at in more detail and discussed in the revised paper.*