Reviewer #1:

We sincerely thank the reviewer for the thorough assessment and valuable recommendations, which have helped us refine the presentation and interpretation of our results. We have carefully considered each comment and incorporated the suggested changes to improve the manuscript. An important change is the addition of a co-author on our paper, Trusha J. Naik, who has performed a sensitivity run with different soil properties, which helped us address a comment of the reviewer.
Below are the reviewer's comments in black, and our replies in blue.

Line 4 –"…model temperatures are consistently too low" implies that this is solely model error, but it is likely a combination of model error and proxy error/bias.

We have rephrased to "lower than reconstructions" instead of "too low", which implied models were wrong.

Line 4 –work."using a previously unpublished geography…", as this will be published

Fixed.

Line 7 – I don't fully understand the sentence starting "GMST varies by up to…". Are these control experiments that just alter CO2, and are otherwise representative of the pre-industrial?

No, these are the Mio_2x and Mio_4x experiments, which are Miocene boundary conditions with either 2 or 4 times PI CO2 concentrations. We have clarified the sentence.

Line 9 – Could you provide an approximate value or range for local warming?

Local warming is around 25°C at the maximum height of the ice sheet (Fig. 7). We have added this in the abstract.

Lines 11/12 – Could you provide an uncertainty range for the ECS estimate, and also for the "modern-based estimates"?

The uncertainty on ECS estimate is 2.9°C (2.5 — 3.3°C, 95% prediction interval), mentioned in L229, and we have now added the uncertainty range in the abstract. We have also added the uncertainty range from PI (Doubling of CO2 from PI: 2.8°C, 2.2 — 3.4°C, 95% prediction interval).

Line 16 – I would suggest a rephrasing of "…cold modelled Miocene climate" as it is a bit confusing given you are speaking about the warm[er than present] Miocene.

We have clarified that our Miocene state is colder than geological data.

Line 18 – The use of "stage" rather than "age" is usually when referencing the rock record only, so I would prefer to see "age" used here.

Line 22 – Similar to above, please use the correct terminology of "age" rather than "epoch" here.

Fixed throughout the paper.

Line 28 – ECS needs to be written out in full on first use, rather than as "equilibrium ECS (ECS)". GMST can also be defined at the end of this line.

Fixed. GMST has already been defined at L19.

Line 32 – The Eocene is an "epoch", not a "period".

Fixed.

Line 33 – Grouping the Pliocene with recent interglacials and the Eocene here misleads the reader, because the Pliocene does have similar CO2 concentration to modern (~400 ppm). This needs to be rephrased. I see value in explicitly stating the Miocene CO2 concentration (with uncertainty range as required) and then relating this to specific instances of near-future estimates (e.g. the estimate for e.g. 2040 under RCP2.6 will be very different to e.g. 2060 under RCP8.5). This is particularly relevant because present-day estimates of ~420 ppm are significantly closer to Pliocene estimates than the ~850 ppm estimate used for the Miocene here.

We agree with the reviewer that mentions about the Pliocene were misleading and weakly connected to the topic mentioned at L33. We have removed references to the Pliocene.

Line 36 – I'd consider citing Burls et al. (2021) when referencing the first MioMIP: Simulating Miocene Warmth: Insights From an Opportunistic Multi-Model Ensemble (MioMIP1) - Burls - 2021 - Paleoceanography and Paleoclimatology - Wiley Online Library.

This reference is provided L35 when discussing previous modelling studies on the Miocene and a clarifying sentence has been added thereafter. "MioMIP1" was not a formal MIP, but an opportunistic comparison, which has led to the formal MioMIP2 (Phase 1 and 2) currently in design.

Line 38 (and similarly in Section 4) – In this paragraph on the 'high latitude paradox', I'd add a more explicit note to say that this is likely a combination of both model error and proxy error/bias (rather than just model error). You could also reference Tindall et al. (2022): CP - The warm winter paradox in the Pliocene northern high latitudes.

We have added some comments about potential seasonal bias on proxy data; a similar point was raised by Reviewer #2. We have also added the Tindal et al. (2022) reference.

Line 45 –"…a previously unpublished paleogeography".

Fixed.

Line 48 –"…constraining modern-day ECS…"?

Fixed.

Line 49 – I would like to see a little more comment on the value that these extra sensitivity studies bring and/or why they are needed.

We have added some extra comments on the importance of large (uncertain) forcing on Miocene climate. We have added a comment on soil properties (colour and consequently albedo) too, as we have added a simulation on that aspect to compare both geographies.

Line 51 – The phrasing of "Langhian or mid-Miocene" is open for erroneous interpretation. Pick one term to use and be consistent throughout the paper.

We have corrected to mid-Miocene throughout the paper and clarified in the Introduction that "Langhian" will be referred as mid-Miocene in the paper.

Line 52 – Is there a model description paper for CESM1.2 that you can cite here?

There is no formal description paper for CESM1.2. Several webpages describing the paper are hosted by NCAR but are no longer publicly available. We have added the extra reference of Neale et al. (2013), which is the closest study of a description of CAM4, the atmospheric model used in our model. We have also added the reference of Zhu et al. (2021), which has also used CESM1.2 (for the LGM).

Table 1 caption – How long were the experiments run for that are not marked by *?

The sentence “All runs have been integrated for at least 3100 years and up to 5000 years when using CAM5 (experiment Mio_noIS_CAM5)” and up to equilibrium” was provided at L96, and we have now added a similar comment as caption of Table 1.

Line 55 – LGM needs to be defined here as it’s the first use.

Fixed.

Line 60 (and 66, 120) – Should “MioMIP2 Phase 1” read “MioMIP1”?

No: MioMIP1 was an opportunistic comparison of models described in Burls et al. (2021). MioMIP2 Phase 1 is a formal approach based on (older) boundary conditions provided by Burls et al. (2021) and which is aimed to be connected to the actual MioMIP2 Phase 2, which will come with updated boundary conditions. Both MioMIP2 Phase 1 and 2 are currently still in design. A clarifiying sentence has been added around L38.

Line 65 – The sentence starting “A comparison of the…” appears incomplete.

Sentence removed (repetition of L59-60).

Line 77 – Is there a citation for the tectonic model of Getech Plc?

The tectonic model of Getech Plc. is private and cannot be shared. However, we have added some extra details regarding the methods (which is similar as Lunt et al., 2016).

Figure 2 – For readers not accustomed to viewing Miocene palaeogeography, some additional shading to mark which area is land and which area is ocean could be beneficial.

Added.

Line 100 – NCAR needs to be defined on first use.

Fixed.

Line 130/131 – Unclear what is meant by the PMOC remaining “a rare phenomenon”, you’ve just said that it’s been simulated multiple times. Do you mean that there is no/little proxy evidence for a PMOC?

Out of the 14 simulations presented by Naik et al. (2025), only 2 simulations have a strong, explicit signal of a PMOC, therefore we can say that PMOC has been in other Miocene simulations but remain rare. Naik et al. (2025) further discuss this, showing that there is almost no geological evidence that a PMOC is still in place during the mid-Miocene. We have added a comment on proxy data here.

Figure 3 and caption – I would like to see individual labels here (a), b) etc.) so that you can add more detail to the caption and refer to specific plots in the main text. Additionally, I’d like to see lat/lon labels for the axes (as in Figure 2) rather than unitless values.

Fixed, and added labels in other figures.

Line 139 – Be cautious how you refer to the different palaeogeographies. Here you refer to “Getech”, but in the previous paragraph it is the experiment name. Be consistent, or for the avoidance of doubt use both (e.g. “the Getech palaeogeography used in Mio_Ctrl”).

Changed to “The Getech paleogeography used in Mio_Ctrl” as we refer to the actual geography here, not a simulation.

Table 2 – Unless the raw GMST values are used elsewhere in the paper then I don’t see the benefit of including both raw GMST and anomaly relative to PI (especially since precipitation is only shown as an anomaly).

Quite often papers use either raw GMST values or anomalies, in particularly when later used by model comparison studies. We therefore decide to keep both to simplify future work.

Line 152 – Is this 14% wetter than Mio_Ctrl or PI? It could be read as either from this sentence alone.

Fixed: It is wetter than Mio_Ctrl and wetter than PI.

Line 162 – Typo on "…largest flora change…"

Fixed.

Line 164 – All models in PlioMIP2 (except COSMOS (Stepanek et al., 2020)) use the static vegetation reconstruction of Salzmann et al. (2008), so this needs to be removed or rephrased to avoid misleading the reader.

We have removed "PlioMIP" and instead replaced it with the reference of Stepanek et al. (2020)

Figure 4 – It is quite difficult to quickly compare these figures as the colours and descriptions are not the same between figures. Could the colour scale of A) be adapted to better aid thiscomparison? Additionally, I would like to see the label for B) reading "mid-Miocene" for consistency with the remainder of the text (to avoid confusion for newcomers to the topic).

The figure taken from Steinthorsdottir et al. (2021) uses biome descriptions that are too different from the biomes of our model to be easily adapted. The figure should be mainly seen as a simple illustration that our model biomes are 1) generally too cold compared to proxy and 2) there is no vegetation at high latitudes. We believe point-by-point comparison is affected by too large uncertainties to be meaningful. We have added the labels A) and B) on the figure.

Line 179 – Add a value to quantify "PI concentrations". Is this 280 ppm? A more specific CMIP6 value?

We willingly refer only to "PI concentrations" here, as the PI value often differs between model (commonly around 280 ppm in paleoclimate settings, but sometimes 284 ppm in certain CMIP6 settings). The PI concentration of our model is provided in Table 1.

Line 181 – Similarly, add value(s) to quantify "high estimates during the MMCO" from Rae et al. (2021).

Added (around 700 - 1100 ppm). Wording in L182-183 was consequently changed to bring more nuance.

Lines 185 and 186 – I feel you need to clarify that you are looking at mid-Miocene in Mio_2x. The reference to the colder Late Miocene is currently a bit confusing.

Clarified at L182.

Lines 187-189 – The sentence "Although there are…" feels a bit unsubstantiated. Can you add more comment here? Expand on the "substantial differences", and why the similarity in continental configuration implies a dominant role in CO2. What about the other non-CO2 factors considered in Burton et al. (2023)?

The sentence was removed. While there are some level of similarity between the Pliocene and Miocene continents, we have concluded that it is too hypothetical to conclude that CO2 is a dominant forcing solely based on the fact that it is dominant at the Pliocene, 10 millions years later, particularly as non-CO2 forcing have substantial regional impact and that CO2 is more transient in the geological record than e.g. continental forcing. We have instead decided to focus only on the comparison between our simulations at 2 and 4 times CO2.

Line 190 – To develop your discussion, I think you can add a more explicit comment that Mio_4x being in agreement with proxy estimates implies that this is likely a good/representative estimate of CO2 for the mid-Miocene.

From the site-by-site comparison provided in Fig.5 and discussed from L200, we know this is actually not the case. While Mio_4x provides a closer match to global temperatures, it strongly overestimates low latitudes, still underestimates high latitudes, and global temperature comparison is not necessarily accurate due to the uneven coverage of the proxy record.

Figure 5 – As with Figure 3, I would like to see lat/lon labels for these axes where appropriate and the addition of a), b) etc. so you can refer to specific plots in the main text. Additionally, an axis label (i.e. Temperature) for the bottom right panel, and clarification of the calculation in the caption (e.g. Mio_2x – PI), would be useful.

Added.

Line 215 – Although PlioMIP and MioMIP have been defined, it is probably best to define "MIP" here.

It has now been defined in the Introduction due to a change of wording following an earlier comment of the reviewer.

Line 220 – Lunt et al. (2010) is arguably the seminal piece on Earth System Sensitivity so should also be cited: Earth system sensitivity inferred from Pliocene modelling and data | Nature Geoscience.

Added.

Line 227 – I would like to see some extra detail added to "We report those simulations…", or remove this sentence and just refer to it in the next sentence as you already do.

Sentence removed (already discussed in the following paragraph).

Lines 229 and 230 – For the reader, it would be helpful to clarify what the CO2 concentration is in ppm for each experiment here. It's quite easy to get lost – one can easily assume that "2x_DblCO2" is a higher concentration than "Dbl_CO2".

Added as well as at L226-227.

Line 236 – I think it is a missed opportunity not to provide a current estimate of modern ECS, e.g. the range presented in IPCC AR6. This is helpful context for the reader, especially those who are new to this area.

We have added the range of the AR6.

Figure 6 – I spent time looking for a black star as is visible in the legend, it needs to be clearer that there is a star for each of the experiments (one in each colour).

Clarified in the caption.

Line 242 – Add a comment on why these sensitivity runs were completed without an Antarctic ice sheet (or add to methods and refer to the relevant section here).

Added. The simulations in this section were performed without an Antarctic ice sheet due to the fact we did not have the Antarctic reconstruction from Getech at the time of the simulation. Because these simulations are expensive and take a long time to run, they were not re-done but simply compared to a mid-Miocene simulation without ice sheet.

Lines 247-249 – This is an important point to make but it would benefit from the additional detail of specific model names. Which CCSM4 version is the warmest? Coldest? For clarity for the

reader, I think it would be better to at least refer to “CCSM4 version” rather than just e.g. “coldest CCSM4”.

Added.

Line 250 – Are you trying to say that CESM1.2 is part of the CESM2 model family here, or that it is a version of CCSM4? It is not clear either way.

We are referring to the family of model nowadays called CESM (1.2, 1.3, 2, 3…), but previously called CCSM. Considering the flexibility of model components and similarities between the models, they are all part of the “CESM” family. We have clarified what we consider as “CESM model family”.

Lines 267-268 – Adding experiment IDs would be useful here, particularly as you refer to Table 2.

Added.

Line 271 – I’m confused that you’re referencing the atmospheric model as an example here, in a section about the impact of the atmospheric model. Is the model affecting global precipitation, or regional?

We rephrased this sentence. There is a coupling between global temperature and global precipitation, which goes beyond the choice of models. However, changing the atmospheric model, in particular in this case, has a strong influence on regions, as it mostly modifies high latitudes cloud parametrizations.

Figures 7 and 8 – As in other figures, please add lat/lon labelling to the axes. I would also like to see the colour bar labels edited to say “surface temperature anomaly” or “total precipitation anomaly” where appropriate, especially as this is not highlighted in the (relatively brief) caption. Readers should also be encouraged to note the difference in scale in the Figure 8 caption (as they are for Figure 7).

Added.

Figure 9 – I think there would be value in labelling which experiment is which dot here, or using different symbols and displaying the experiment IDs in a legend. This would add more depth to the analysis and aid further discussion.

We have added experiment numbers and captions. The dots are reported from Table 2.

Overall comment – I would like to see a bit more discussion here. What are the implications of the choice of atmospheric model being important (e.g. for MioMIP)?

We have added a concluding remark on the choice of atmospheric model as to answer the question raised at the beginning of the section. Quite often there is no “choice” of atmospheric model: modelling centres handle the experimental design with what is possible time-wise and resource-wise. This section highlights that changes in atmospheric parametrizations, in particular in relation to cloud (and so cloud feedbacks) can have substantial impact on the climate, and can easily justify that several models with apparent similarities be together within a MIP.

Line 297 – By “at its highest point”, do you mean highest point in elevation, or greatest temperature anomaly? This needs to be clearer.

Fixed “highest altitude”.

Overall comment – Again, a bit more discussion would be valuable. E.g. could this be a lesson going into MioMIP or future Miocene modelling efforts?

We consider the section to provide sufficient detail demonstrating that the solar forcing is small (approximately 4% of the CO_2 forcing). Moreover, solar forcing is comparatively well constrained

and is not expected to change substantially in future modelling experiments. For these reasons, we expect it to remain negligible in the context of the present study.

Line 317 – Like elsewhere, I'd phrase as "…new and previously unpublished paleogeography".

Fixed.

Line 319 – I think you can go further than just saying "…consequently the scale of the warming" here. This underlies the possibility of using the Miocene as an analogue, and these uncertainties are rightly mentioned alongside a suggestion of analogy. Even if temperature is shown to be similar to future projections, if CO2 is seen to be non-dominant in the Miocene, can we really call it an analogue? What about the AMOC shutdown? There is some discussion around the need for critical thinking in this area in Burton et al. (2023: CP - On the climatic influence of CO2 forcing in the Pliocene), Oldeman et al. (in review: A Framework for Assessing Analogy between Past and Future Climates by Arthur M. Oldeman, Lauren Burton, Michiel L. J. Baatsen, Henk Dijkstra, Anna von der Heydt, Aisling Dolan, Alan M. Haywood, Daniel Hill, Julia Tindall :: SSRN), and Burton et al. (2025: An assessment of the Pliocene as an analogue for our warmer future - ScienceDirect).

We have added comments on this aspect, and we thank the reviewer for the references. We believe the term "analog" to compare global climates is in fact a poor term, as supported by Oldeman et al. (in review), which suggest "partial analogy" to be more appropriate. As the reviewer is pointing out, aspects such as the AMOC shutting down, uncertainties on certain forcing (CO2 and continental), and in particular ECS being close to modern estimates while global and regional climate is so different, show that "analogue" is a poor term to compare the Miocene to modern climate. We have rephrased some aspects of the conclusion to emphasise that the Miocene is only a partial analog (which is not a weakness of the paleoclimate, as mentioned by Oldeman et al.).

Lines 351-352 and 355-357 – Representing mid-Miocene climate… assuming that the high values indicated by proxies are correct. There is likely a middle ground here where both models and proxies are biased. Turning up the CO2 to unevidenced levels in the models is not the right approach!

While we agree that proxies might be wrong as well as models, the CO2 concentrations mentioned here and tested in our paper are not unevidenced, but are supported by multiple proxy studies. We have however added a comment on the middle ground which should exist, where both proxies and models are wrong altogether. We have already commented in our paper that turning up CO2 is not a right approach due to the fact that it overshoots low latitude temperatures (Section 4).

Appendix A: Impact of ice sheet on heat transport and radiation
Reference to Appendix A comes after reference to Appendix B in the main text, so they need to be switched around (i.e. Appendix A becomes Appendix B, and B becomes A).

Only one appendix is now left.

Line 507 – Typo on "Arctic"
.
Line 508 –
"increased" should read "increase"
.
Appendix B: Additional figures
The appendix only contains one additional figure, so I would like to see the header rephrased and perhaps a sentence or two describing the figure.

The figure has been moved to the main text.

Reviewer #2:

We appreciate the reviewer's detailed and constructive feedback, which has contributed to enhancing the clarity and robustness of our study. In response, we have revised the manuscript to address all points raised and clarified the relevant sections accordingly. Below are the reviewer's comments in black, and our replies in blue.

Paleogeographic Reconstruction:
The study relies on an unpublished paleogeography from Getech Plc., which is compared with the Burls et al. (2021) reconstruction. While this offers a valuable comparison, the manuscript provides insufficient details on the paleogeographic reconstructions, and does not compare with several recent published mid-Miocene paleogeographic reconstructions. As the paleogeography is the important boundary condition that fundamentally shapes the simulated climate.

We have added some comparisons to other published geographies (Frigola et al., 2018; He et al., 2021), notably emphasising differences in the Tethys, the Fram Strait and the Panama seaway. We have moved a paragraph describing differences in paleogeographies which was in the results to the Methods section as well.

Model-Data Mismatch:
The manuscript highlights the persistent cold bias in high latitudes. It is advisable to consider adding a discussion on "seasonal deviation" of the proxy. And the discussion could be strengthened by linking this to broader challenges in paleoclimate modeling (e.g., cloud feedbacks, ocean heat transport, vegetation-atmosphere interactions).
Addition, a key concern is that the model might be over-reliant on CO_2 forcing to achieve a warmer climate, potentially at the expense of other, less constrained factors that are known to be important for warm paleoclimates, such as the role of ocean–ice interactions and their dependence on model dynamics.

We have added a comment on potential seasonal bias from the proxy record. As also mentioned by Reviewer #1, we have added some emphasis on the large error that could come from the proxy record, and not just models. Some model challenges are already discussed in our paper, for instance dynamic vegetation or polar amplification factors, which are known issues of the modelling world. We have added an extra comment discussing the lack of interactive ice sheet, which can impact high latitude temperatures.

The simulated AMOC:
The model simulation suggests a collapse of the Atlantic Meridional Overturning Circulation (AMOC) during the middle Miocene. However, the available geological evidence does not support a complete shutdown of the AMOC at that time.

Indeed, Naik et al. (2025) shows that proxy records support more an AMOC state and a collapse of a PMOC state earlier than the mid-Miocene. However, several models have managed to maintain a PMOC state under similar boundary conditions, which highlight the critical need of investigating these models (and our research) as to under what conditions this is possible and what could it indicate for future climate change. We did not discuss the collapse of the AMOC further in our paper as it is kept for a future study which is currently being worked on. Some of these comments were added around L155 and L224.

Antarctic Ice Sheet Impact:
The finding that the inclusion of an Antarctic ice sheet leads to global warming is intriguing and counterintuitive. While the authors briefly discuss this, the mechanism remains speculative. Provide the relative contribution percentages of each feedback item (cloud, water vapor, albedo, temperature) to the 1.4 °C increase. This could be a significant contribution if better explained

While the suggestion of the reviewer is definitely interesting, it represents a tremendous amount of work which is a study on its own. Splitting the contribution of each individual feedback can be done with approximate partial radiative perturbation (APRP) or online PRP, a process that would require the development of such module, its implementation in the model, re-running the simulations with usually a running time which is doubled for online PRP.