Reviewer 1

General comments

This paper uses the results from PlioMIP2 to examine the NAO and North Atlantic circulation during the late Pliocene. The topic is of interest and importance, with both significance to paleo and future climate. Therefore, there is definitely a need for this research. It can help to fill in the research gaps on North Atlantic variability during the LP, plus providing a better understanding of the behaviour of the NAO is vital to improving modelling work focusing on future projections.

Considering this importance, I think the paper could do a better job of communicating its findings. The work conducted is extremely thorough and methodically robust, yet the results are hard to extract, and the conclusions perhaps unclear. I think with some restructuring and a better focus on answering the hypothesis the paper sets out to do, it would greatly improved.

The authors thank Anonymous Referee #1 for their comments which will improve this manuscript. Below we detail how we will implement the suggested changes into the manuscript.

Introduction

I feel this section could be improved by condensing 1.1 and 1.2, and with a slight restructuring. There is a fair amount of repetition on the importance of studying the NAO, which I agree with, but it is currently scattered throughout the text. It may have more impact if there was a dedicated paragraph detailing this. On that note, section 1.3 is interesting and well written. I would condense the text prior to this, so you arrive at this section faster. This is what makes the study unique and interesting so let this be the focus.

The introduction will be shortened by removing some details that are repeated in the methods. The section of the introduction about the uncertainty in modelling the NAO will be moved into the discussion section.

I am confused by lines 27 to 40. It states "studies of the LP", which studies? Are these the studies then referenced in lines 29-30, or are these studies that have only focused on the long term state? But then Pontes et al is referenced again when discussing high frequency. If the aim of these paragraphs is to identify the research gap or where current literature has not yet gone, perhaps be clearer on this. Lines 40-44 are really clear with the delivery of the aims, so tidying up the paragraphs before so they lead into this could really help.

The purpose of these paragraphs is to introduce how modelling the Pliocene contributes to improved understanding of climate (both past and future). It also introduces the studies that have been completed that examine higher frequency variability. The wording of these paragraphs will be adapted to improve clarity.

Pontes et al., 2022 is referenced twice, once for the change in Hadley cell circulation (mean state) and again for the reduction in ENSO variability (variability). This has been clarified in the text by specifying that the literature referenced are based on mean state.

Results and Interpretation

I understand that this section is both a combined results and discussion, but I found it hard to draw out the results due to the constant discussion elements. I think this section could benefit from some restructuring.

Section 3 – results and interpretation has now been split into two sections including a dedicated results chapter and a sperate discussion chapter. This should improve the clarity of the results found in this study.

Firstly, the results that are presented are quite vague. The wording is often "general agreement," "larger changes", "diverse changes", "more positive on average," but these are more descriptive than concrete results. Quantifying these changes would make it much clearer what the study has found and why it is important. Statistics to actually demonstrate the range of what diverse means, or how agreed results are are needed. The maps and figures 4-5 show the changes, but quantities are hard to interpret from the figures alone.

The results section will be adapted to be more quantitative in the results presenting the values shown in the figures to improve clarity.

Secondly, I think the flow would be improved by establishing the results first, and interpreting them after. The continued sentences with very speculative wording like "which suggests," "indicates", "may also be influenced," "likely due", not only makes it harder to find, absorb and remember the results, but it makes the reader continually question the results presented. Whilst interpretation and discussion is vitally important, and I don't want it be removed, I wish that the findings had been presented in a clearer way, before the issues associated with them were dissected.

The results and discission section will be split into two sections. Along with a more quantitative description of the results, this should improve the clarity of the results presented.

In this vein, I think a dedicated limitations paragraph is needed, as opposed to having single sentences spread out. This could delve properly into the issues working with higher frequency variability modes and wind characteristics, with just monthly data from PlioMIP simulations. This is touched on briefly in line 234, but is important to address as it must have a big impact on the result, or what we can learn from it. A limitations section could also cover the other factors that must have an impact on the findings but are only sporadically mentioned (e.g line 193 vegetation, line 199 orography, line 248 East Atlantic pattern). By having a dedicated limitations, it would reduce the amount of doubt that's continually cast on the findings, and instead present it more in a light of what we can be done better in future. This would also help improve the conclusion- see below.

There will be a limitations paragraph in the discussion section of the revised manuscript which will cover the issues currently spread throughout section 3.

Section 3.4 Impact of NAO, is interesting. Evaluating the impact on the climate is, I feel, at the core of the significance of the study and should be thoroughly explored. Quantifying values would aide the section, "Northern Europe has increased precipitation whereas Southern Europe ..." how much? Same can be done for temperature. From figures 6/7/8 it is hard to define what amount the colours respond to.

In the revised manuscript, the results of this section will be presented more clearly by stating values as suggested. The colour bars on the Temperature plots in Figs. 7 and 8 will be changed to improve clarity of the figures.

Section 3.5 I am slightly confused by. No data comparison has been done, although possible datasets are mentioned, but the problems associated with data comparison are discussed and future possibilities. Also data comparisons are mentioned in the previous sections to, so there is somewhat repetition. Therefore, i struggle to find the outcome of this section. If it is not possible to do a data comparison, then this is something to be discussed in limitations, along with what is needed to in the future. Or if the authors want to do a data comparison, however howsmall, I think it would benefit the study. This perhaps goes back to my earlier comment on providing quantities, definite numbers may be easier to evaluate against proxy reconstructions then just descriptions.

In the revised manuscript, the section will be rewritten to improve clarity. It is currently not possible to provide a quantitative comparison as the data available are sparce and not dated to the Late Pliocene orbital window used in PlioMIP2. Therefore, this section will also include a discussion of the potential for future data model comparison and what proxy data would be useful to complete this analysis in the future.

Conclusion

This section is concise and well written, but considering the sporadic nature of the results, I was hoping that the conclusion would be more affirmative in what was achieved. Perhaps with a less descriptive and more quantative results section, the conclusion can naturally improve, with a clearer message in what was achieved. Parts of this section, e.g. line 325 would work better in a separate limitations section.

The revised manuscript will contain a separate results and discussion section to improve the communication of the main findings of the study which should improve the conclusions. Parts of the conclusion will be placed into the new limitations section in the discussion to improve the flow of the conclusions section.

Minor Comments

• Line 18: "and palaeoenvironmental reconstructions using proxy data" Unclear phrasing, do you mean climate is focus of both the modelling and the proxies?

This will be rewritten to the following text to improve clarity.

“The Late Pliocene (specifically Marine Isotope Stage KM5c, 3.205 Ma) climate has been investigated through both modelling and proxy reconstructions. It is the focus of the second phase of the Pliocene Model Intercomparison Project (PlioMIP2) (Haywood et al., 2020, 2024), and several palaeoenvironmental reconstructions using proxy data (e.g. Foley and Dowsett, 2019; McClymont et al., 2020).”

• Line 19: change to "comprises of outputs from 17"

This change has been made as suggested.

• Line 20: "consistent manner" wording is too vague, do you mean with same experiment setup?

To improve the clarity of this, the wording will be changed to: “PlioMIP2 comprises outputs from 17 atmosphere-ocean coupled climate models with a similar experimental set up following the Pliocene Research, Interpretation and Synoptic Mapping Version 4 (PRISM4) boundary conditions (Haywood et al., 2020).”

• Line 21: "Late Pliocene experiment, CO2 concentrations were set to 400 ppmv and the Antarctic and Greenland ice sheets were reduced in size compared to modern." You state this later in the methods, i think keep it there and remove it from the introduction as it doesn’t quite fit.

This section will be removed from the introduction.

• Line 24: "The PlioMIP2 multi-model mean (MMM) simulates global annual mean surface" Is this referring to MMM of LP core experiments? Specify.

This will be re-written to: *“The PlioMIP2 multi-model mean (MMM) simulates global annual mean surface air temperatures of 3.2°C above pre-industrial and a global annual precipitation 7% above than pre-industrial in the Late Pliocene (Haywood et al., 2020).”*

• Line35: "report a reduction" Unclear statement. A reduction in LP? A reduction compared across the models? Compared to what?

This statement will be reworded to improve clarity. *“Pontes et al. (2022) and Oldeman et al. (2021) report reduced El-Niño Southern Oscillation (ENSO) variability in the Late Pliocene compared to the pre-industrial across the PlioMIP2 ensemble.”*

• Line 36: "Changes in circulation patterns", this is unclear, does this refer to changes in ENSO? Or Atlantic?

This has been re-written to *“Changes in atmospheric circulation patterns”* to improve clarity.

• Line 47: " over the North Atlantic associated with", change to "which is associated"

This change has been made as suggested.

• Line 52: "closer to Ireland" and "in summer the westerly winds sit equatorward". These sentences feel incomplete or arbitrary- either explain or remove.

This section has been re-phrased to: *“In the Northern Hemisphere summer, the location of the high pressure centre shifts northeast. Like the winter NAO, the summer NAO is related to the westerly winds where in summer the westerly winds sit equatorward during a negative phase and poleward in a positive phase (Dong et al., 2013).”*

• Line 54: "a strong influence on European climate and hence on agriculture" change to "on European climate, which affects agriculture and ..."

This change will be made as suggested.

• Line 59: "present day climate" maybe would be better referred to as "modern climate"

This change will be made as suggested.

• Line 61: "with reanalysis data showing a stronger relation to jet stream" this is unclear, a stronger relation of what exactly?

This line refers to the relationship between trends in European precipitation and the speed of the jet stream. In reanalysis data, the relationship between changes in European precipitation and the change in jet stream strength is stronger than the relationship seen in climate model simulations, implying that models might not capture these changes. This section of the introduction will be moved into the discussion section (to condense the introduction as suggested above) and this will be discussed as a potential limitation of the climate model outputs. The wording will be changed to improve clarity.

• Line 113: "PlioMIP2 data" very small thing, but data is maybe more associated with proxies, so maybe use model output instead?

This change will be made as suggested.

• Line 140: "fully separate out the orography and ice sheets". I can't find if the ice sheets are interactive or static, perhaps add this in.

The methods section will be revised to include reference to the static ice sheets used the PlioMIP2 simulations. It will also include more details on the models that use static vegetation vs the model that uses dynamic vegetation. The impact of not being able to sperate out the orography in ice sheet regions from the ice sheets will also be listed in the limitation section of the revised manuscript.

• Line 163: "pressure that is too high" and "too low". This wording is misleading, too high/low implies it's wrong, and if this is a comparison with reanlysis- which you have previously stated cannot be fully trusted, then there isn't necessarily a right or wrong.

The wording will be fixed to remove the implication that the pre-industrial simulations are wrong and instead comment on the differences in terms of a potential bias and to help explain differences across the model ensemble.

• Line 167: Why do we already jump to figure 6 and 7? what happened to 2,3,4.

These figures were supposed to reference to Figures S6 and S7 in the supplementary information which will be fixed in the text.

• Line 173: "opposite to that projected by higher (abrupt 2×,4× and 8× CO2)" I don't quite understand the importance of this, I can't find if this simulation is discussed again or why it matters in the LP context.

This was included to describe the type of experiments but "(abrupt 2×,4× and 8× CO2)" has been removed to avoid confusion with Late Pliocene simulations.

• Line 183: "change in SLP is diverse" vague wording, what does this mean? Can we have some quantification of what diverse is?

Late Pliocene simulations. The aim of this statement is to address that the total change in sea level pressure is not consistent across the three models used for the forcing factorisation analysis with COSMOS and CCSM4-UoT showing opposite responses. The text has been adapted to improve clarity. The new text will read:

"In the winter months, the total change in SLP is inconsistent across the three models used, with CCSM4-UoT and COSMOS having opposite responses (Fig 2)"

• Line 197: "Changing the Greenland ice sheet impacts North Atlantic circulation. By implementing Late Pliocene ice sheets, the orography of Greenland is changed" Consider rewording this. "By implementing Late Pliocene ice sheets, the orography of Greenland is changed, which in turn affects North Atlantic circulation"

This change has been made as suggested.

• Line 206: "more positive on average (Fig 4). However, there is large model spread" This would really be improved with quantitative support. How much? What range?

In the revised manuscript, the results section will contain more quantitative descriptions of the results. For example:

"Across the PlioMIP2 ensemble, the NAO is more positive in the Late Pliocene simulations compared to the pre-industrial simulations with the multi-model mean being 2.9 hPa stronger. There is a large model spread in the response to the NAO with a range from – 5.4 to 14.5 hPa. Most models (10 out of 17) display a weak positive difference (less than 2.5 hPa) with two models displaying a more negative average NAO index in the Late Pliocene and 5 models displaying a more positive difference (between 5.6 and 14.5 hPa)"

• Line 231: (Fig. 5 Missing bracket.

This change has been made as suggested.

• Line 314: missing comma after ensemble

This change has been made as suggested.

• Line 322: "data set are likely within the range of internal variability of the models" Does this mean none of the NAO results can be trusted? Perhaps consider rewording- or address properly in limitations to avoid confusion.

Due to the limitations of the data with only one 100 year period for each model simulation available it is difficult to assess the impact of internal variability on the change in NAO variability between the two time periods. Furthermore, longer modes of variability (such as the AMO with a period of ~60-90 years) could also impact the results. These factors will be discussed in the limitation section.

• Figure 5: "Change to the North Atlantic" Please specify change from what to what.

The figure caption will be updated to improve clarity by including the below text

"Difference in the North Atlantic zonal wind speed and latitude of maximum wind speed between the Late Pliocene and pre-industrial simulations for each model."

• Figure 8: Panel showing Residual, it is very hard to distinguish the different red shades and therefore differences.

The colour map will be changed in the revised version of Figure 8 to be clearer in the different shades of red. The differences will also be presented in the text in the new dedicated results section

• Figure 9: "The change in the mean" Please specify if this is LP-PI difference. I think the values that have been added to this figure would be far easier to read as a table.

The values will be added to a table in the supplementary information which will be referenced in the caption of Figure 9 and in the text of the results section.

- **Figures 1-3 and 6-8: The boundary lines on the continent are way too thick and make it the maps very busy, especially around Canadian and Norwegian coasts.**

In the revised manuscript, the figures with maps will use thinner lines for the continent boundaries.