Reviewer 2

General assessment:

I find the research relevant, both for paleoclimate research and for interpreting paleoclimate in the larger picture of climate change. The presentation, however, can still be improved. The figures are of high quality throughout the manuscript and supplement, which is very laudable. My main concern is the organization of the information and the text flow: over large parts of the manuscript, the red thread is missing. On the other hand, I also think that the authors are in parts underselling their work. It could be highlighted more clearly that the reduced ensemble is quite valid towards providing results that are representative for the wider PlioMIP2 framework. See my comments and suggestions below.

The authors thank Anonymous Referee #2 for their comments will improve this manuscript. Below we detail how we will implement changes to the manuscript.

====

MAJOR COMMENTS

====

M1. Structure and red thread

1. Introduction. Most of the information is relevant, but it should be brought into a more logical order, and three separate sections may not be needed. Please order the information to show (1) what we already know, (2) what we still would like to learn, (3) where current and previous problems lie, and (4) what this study contributes to move us further.

The introduction will be condensed with the information around uncertainty in modelling the NAO moved into the new discussion section. The three sections are retained but section 1.1 and 1.2 have been shortened.

2. Sect. 3. Results and interpretation are tightly interwoven. This may be intentional given the heading of Sect. 3, but I found it difficult to separate the findings of this study from their interpretation and from the link to previous work. There are enough results to justify presenting your own findings and their interpretation in separate sections. I suggest revising Sect. 3 accordingly.

The results and discussion will be presented in two different sections in the revised manuscript.

3. Discussion. A clear discussion section is missing (see above).

Section 3 – results and interpretation will be split into two sections including a dedicated results chapter and a sperate discussion chapter.

4. Abstract. Please embed the utility of the Pliocene NAO for future warming better into the reading flow, e.g. by presenting the study as a critical evaluation of the degree of analogy.

The below text will be included in the abstract to comment on the Pliocene NAO as an analogue.

"Using forcing factorisation simulations, the differences in sea level pressure are found to be driven by non-CO2 boundary conditions such as ice sheets and orography, which has implications for using the Late Pliocene as an analogue for future North Atlantic variability."

The main purpose of this paper is to present potential implications of changes in modelled variability rather than an assessment of the Late Pliocene as an analogue for future variability. To consider how the Late Pliocene performs as an analogue, simulations for the future would be needed which are beyond the scope of this paper. This will be clarified in the text to better communicate this.

M2. Methods

5. Orography and pressure. Eoi400 differs in orography from the modern state. Depending on the model and on how the modelling groups processed their pressure output, this may cause offsets between E280 and Eoi400. I do not expect this to affect the NAO results, but some pressure anomaly plots may differ between models not only because of model dynamics, but also because of different treatments of model output. Please verify that each model's area-weighted global mean pressure is the same in its respective E280 and Eoi400 simulations. If not, I think a standard approach is to correct the Eoi400 mean to be the same as the E280 mean.

The reviewer is correct that some differences will arise in the pressure anomaly plots due to model treatment of pressure. However, surface pressure data are not available for the full ensemble. Due to the temperature dependency of sea level pressure this correction would need to be done with surface pressure. Because of the lack of data, we have chosen to not apply this correction to only some of the models but have noted this as a potential uncertainty in the manuscript.

6. Lines 135ff (index vs. EOF). Considering EOFs in addition to the index analysis is a good idea. Please state more clearly why both routes are followed and which one is better suited to your work and to the Pliocene.

As the NAO index has two common definitions, the EOF and the station based, we tested both to ensure that the results were independent of the choice of method. One potential issue with the station-based method would be if the pressure centres that form the NAO moved in the Late Pliocene simulation compared to the pre-industrial which may impact the NAO index. To test this, we compared the two methods (Figure S1 and S2) and found they produced similar results. As the station-based method is more direct related to the physical process rather than a statistical product as the EOF analysis, we chose that index definition. This will be clarified in the methods section by including the following text:

"This station-based metric provided a similar result to using the leading empirical orthogonal function (EOF1) (Figs. S1 and S2). The station-based method is followed here as it provides a more physical interpretation in contrast to the EOF statistical method."

7. Lines 217ff (internal variability). The NAO has a variability in the order of ~5 years, and the PlioMIP2 modelling groups should mostly have submitted at least 100 model years of data, so internal variability can be tested. This is the standard for equilibrium runs in PlioMIP. If, instead, you want to address with your statement the topics transient forcing, state dependence (e.g. ENSO or AMO), stationarity, autocorrelation, or drift, then please make this explicit. These aspects could also be discussed more explicitly in the revised manuscript.

As there is only one simulation (of 100 years) from each model, longer modes of variability (such as the AMO with a period of ~60-80 years) may introduce some uncertainties in the analysis. It is

common to assess the role of internal variably by comparing across ensemble members with sea level pressure having large uncertainty in the forced response. Therefore 100 years of 1 model may not be sufficient to capture the true range of the variability within the model. The section will be rewritten to include mention of how the change between the pre-industrial and the Pliocene compared to the standard deviation of the pre-industrial NAO. For example, COSMOS has a pressure gradient difference of 13.5hPa with a PI standard deviation of 8hPa. However, this method is not sufficient for providing an assessment of the robustness of the change in the variability of the NAO. This will be discussed in the revised manuscript.

M3. Proxy and vegetation discussion (Sect. 3.5)

8. Tone of the proxy-data discussion. The discussion reads rather pessimistic and leaves open whether any useful path exists. One could rather highlight here what is already possible and where to go in the future - unless you are really that pessimistic, but then I do not understand the various suggestions that you present where you seem confident that model-data comparison is generally useful.

In the revised manuscript, this section will be rewritten with a focus on what data exists, and what data would be needed to perform a data-model comparison.

9. Fixed vs. dynamic vegetation. I think that this aspect of the discussion could be extended. Since you highlight that vegetation has some impact, I as a reader would be interested in your opinion on the use of fixed versus dynamic vegetation simulations and on how each furthers our understanding in different ways.

This topic will be expanded in the revised manuscript to comment of the impacts of dynamic vegetation on climate patterns and variability. It is difficult to draw conclusions on this from the PlioMIP2 ensemble as only one model has dynamic vegetation, so it in uncertain if the changes are due to dynamic vegetation or if this fits within the model spread.

10. Seasonality and proxy resolution. Please discuss whether high temporal resolution proxy data is really necessary to evaluate model performance on seasonal time scales. Indirect approaches may exist, e.g. deriving from plant types whether springs were wet and summers dry rather than the reverse. A more detailed discussion would increase the impact of your study.

Seasonal proxy data would still provide a 'mean state' view on the climate. The aim of this study it to look towards higher frequency variability and how/if changes in the NAO impact distributions of precipitation in the Late Pliocene. To validate these higher-frequency changes then high temporal resolution proxy data would be needed. The revised Section 3.5 will clarify this but also include some discussion of seasonality and how the changes in the NAO may impact that.

11. Verifiability. Please check that statements are verifiable. The last sentence on p. 16 is supported neither by a reference nor by numbers or findings from your own study.

The revised manuscript will be clear on where statements are from this study or from other findings.

12. Structure of Sect. 3.5. Please make clear that the vegetation discussion ends on p. 16 and that other proxies follow. This only becomes apparent later and is confusing.

In the revised manuscript this section will be rewritten for improved clarity.

13. Line 294. This statement is not specific to paleoclimate, since it also holds for the present climate. I suggest stating that paleoclimate provides model-independent verification data from periods for which models have not been regularly tuned. Models are commonly tuned to modern climate, which makes verification there difficult.

This statement will be expanded on and highlight the benefits of using paleoclimate proxy data to verify model performance.

=====

MINOR COMMENTS

=====

<u>m1. Terminology and consistency</u>

14. Model names. Please use model names consistently across manuscript and supplement, e.g. CCSM4-UoT vs. CCSM4-UofT.

The revised manuscript will be consistent with the use of CCSM4-UoT.

15. Eoi400/E280 vs. Late Pliocene/preindustrial. In most of the main text you refer to Eoi400 as Late Pliocene, but not always. Do you intend Eoi400 and E280 as terms for the simulations, and "Late Pliocene" and "preindustrial" for the periods more generally? If so, please apply this distinction consistently.

The revised manuscript will remove mention of Eoi400 and E280 and instead refer to the results as Late Pliocene simulation and pre-industrial simulation. This is to improve the text flow for those unfamiliar with the PlioMIP naming conventions and to be consistent with future phases of PlioMIP.

16. Fig. 9 / "NAO index". The caption says the NAO index is annotated, but pressure values are shown next to the model names. Strictly speaking, this is the pressure gradient, not the NAO index. Please check the whole manuscript for consistent use of this terminology.

As this study uses the station-based method, the NAO index has units of hPa because the index is defined as the pressure difference.

17. Line 30. Should sea ice be listed within a group of large-scale circulation patterns, or should the two be separated more clearly?

Sea ice will be listed separately and include more references to potential impacts of the sea ice on mid-latitude climate.

18. Line 245. The summer zonal wind speeds are described as "more equatorward". I think you mean that the latitude of the jet (or of the maximum) is more equatorward, which corresponds to a negative Δlatitude.

The zonal wind speeds are referenced as being more equatorward in the Late Pliocene, the revised manuscript will be more consistent with term ‘jet’ and ‘westerly winds’.

<u>m2. Methods and framing details</u>

19. Excluded models. Not all models are shown in all figures. Please state for the various figures the reason for exclusion (I assume data availability).

Two models are excluded from figure 5 due to data availability, this will be stated in the methods section of the revised manuscript.

20. Line 86. I found it difficult to locate, in previous publications, how the Rocky Mountains differ between PlioMIP2 and PlioMIP1. Since the authors emphasize their potential impact, please indicate clearly where this information can be found.

Both PlioMIP1 and PlioMIP2 contain Rocky Mountains that are closer to their modern-day height based on PRISM3D and PRISM4 (Haywood et al., 2010 and Dowsett et al., 2016) boundary conditions. However, in PRISM2, the Rocky Mountains were reduced in height by 50% (Dowsett et al., 1999). The previous study on the NAO, Hill et al., (2011), uses the PRISM2 boundary conditions which as they highlighted in their study could impact the results. The revised manuscript will include references to these publications.

21. Line 110. The meaning of "initial assessment" is unclear to me. The PlioMIP2 model output is already well established, so do you plan more detailed work on precipitation shifts in the future? Please clarify.

'Initial assessment' was used as to fully understand the changes in the distribution of precipitation; more analysis would be needed which is beyond the focus of this paper which mainly looks at the NAO. This will be rewritten to improve clarity.

22. Line 145. Consider adding "despite coarse model grids" at the end of the sentence to highlight the utility of your approach.

This will be added to the revised manuscript as suggested.

m3. Results and figures

23. Fig. 4 (highlighting the validity of your reduced ensemble). For some analyses you use a reduced ensemble because data from other PlioMIP2 models are missing. Nevertheless, Fig. 4 shows in my opinion that your reduced ensemble actually brackets the mean Eoi400 E280 NAO anomaly of the full PlioMIP2 ensemble quite well, best for DJF but also for JJA. For the standard deviation, the coverage of the extremes is weaker but still acceptable. I suggest highlighting this as a strength, which would also preclude potential doubts about robustness regarding the reduced ensemble size.

In the revised manuscript it will be noted that the three models used in the forcing factorisation experiment are representative of the wider range of sea level pressure difference across the whole ensemble.

24. Lines 210 and 216. Please emphasize once more that these statements are based on your results rather than on a different publication.

This will be clarified in the revised manuscript as the results and discussion are separated into two sections.

25. Line 225. The question of variability changes and their impact on European climate should be better linked and discussed. Here it stands somewhat disconnected.

The revised manuscript will contain more discussion on how a change in NAO variability would impact European climate. In the new discussion section, it will be integrated so it is less disconnected.

26. Line 266. Does "still" refer to the Pliocene here? Please clarify.

In line 266 "still" was intended to refer to the Late Pliocene. This will be clarified in the revised manuscript.

27. Fig. 9. Please consider also identifying the wettest month, to see whether the models diverge strongly. Shifts in precipitation seasonality may be as relevant as changes in its amount. You could refer to Fig. S14 here or include directly in this figure information on the wettest month if meaningful.

In the revised manuscript, reference to Fig. S14 and the variation of the wettest month between models will be noted.

28. Fig. 1. Please state whether panels d and h show the same region as the spatial plots or a wider zonal mean. If possible, also show the NAO index boxes in the other maps; this would make the results description in lines 170ff easier to follow.

In Fig 1, panels d and h are showing the zonal mean of the spatial plots. This will be clarified in the figure caption. The boxes depicting the NAO index region will be added onto Fig 1. C and g.

m4. References

29. Line 18. Please check whether citing Haywood et al. (2024) in the PlioMIP2 context is intended.

The Haywood et al., (2024) reference will be removed from this line.

30. Lines 23-24. These statements should be supported by a reference. Otto-Bliesner et al. (2017) comes to mind. Recent vegetation modelling work may also be useful, or you could refer to the boundary conditions published within PlioMIP2/PRISM4.

The revised manuscript will contain references to the boundary conditions from PRISM4 (Dowsett et al., (2016)) and the vegetation reconstruction (Salzmann et al., (2008)) and the impacts of some of these changes on the climate (Otto-Bliesner et al. (2017)).

31. Line 42ff. A reference seems to be missing.

The revised manuscript will now reference studies of variability in the late Pliocene. The studies that will be reference are Oldeman et al., (2024) and Canal-Solis et al., (2025).

====

TECHNICAL COMMENTS

====

All technical comments will be incorporated as suggested.

General

Commas. Additional commas would clarify the text flow, in particular after introductory phrases and in complex sentences. Examples: line 290 ("Although negative on average, there are no ..."), line 108 ("To achieve this, data ..."), line 129 ("of the sample period, meaning"). Please also check for misplaced commas, e.g. line 316: "In winter months, on average the, NAO".- Repetition. Please remove repeated words in close proximity, e.g.

"increased" in lines 4 and 5.- Supplement references. Please ensure that Figs. S6, S7, S9, and S14 are referenced in the manuscript.

By line

Line 12: fix the grammar ("... but that these differences are may not be related to ...").

Line 19: fix "from of".

Line 23: typo "Baring".

Line 25: fix "above than".

Line 55: should this be "predicting the future behaviour"?-

Line 59: "it is known that" before "the wintertime" would clarify the formulation.

Lines 61/62: the wide bracket can be removed; the text reads well without it.

Line 63 (also line 125): "Models'".

- Line 66: fix the circular formulation.-

Line 142: "the the".

Line 167: the reference should be to S6 and S7.

Line 176: add a reference to subfigures d and h.

Line 231: add the closing bracket.

Line 236: "display" in plural.-

Line 277: "Pliocene NAO on European climate,".

Lines 284/285: "In northern Europe ... in northern Europe" is repetitive.

Line 309: fix the grammar. I am also unsure what information shall be conveyed here.

Figures

These comments will be included in the revised manuscript as suggested.

Fig. 6: the subfigure references in the caption are misleading. It should maybe read:(a-c) pre-industrial simulation (E280), (d-f) Late Pliocene simulation (Eoi400), and (gi) the difference between them (Eoi400-E280). Note the missing closing bracket. The sameapplies to Fig. 7.-

Fig. S2: it seems compressed. Using the full text width would give the y-axes much morespace.

Fig. S14: use "crosses" instead of "crossed", and consider referring to the domain definition for Northern and Southern Europe.

Reference

Otto-Bliesner, B. L., A. Jahn, R. Feng, E. C. Brady, A. Hu, and M. Löfverström (2017), Amplified North Atlantic warming in the late Pliocene by changes in Arctic gateways, Geophys. Res. Lett., 44, 957-964, doi:10.1002/2016GL071805

References:

Canal-Solis, K. V., Tindall, J. C., Oldeman, A. M., Haywood, A. M., Abe-Ouchi, A., Baatsen, M. L., Chan, W.-L., Feng, R., Hunter, S. J., Li, X., Stepanek, C., and Zhang, Z.: Asymmetric Pacific variability in the Pliocene: An unchanged PDO relative to a suppressed ENSO, Global and Planetary Change, 253, 104 932, https://doi.org/10.1016/j.gloplacha.2025.104932, 2025

Dowsett, Harry J., et al. "Middle Pliocene paleoenvironmental reconstruction: PRISM2." *US Geological Survey open file report* 99 (1999): 535.

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., Pound, M., Salzmann, U., Robinson, M., Chandler, M., Foley, K., and Haywood, A.: The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction, Clim. Past, 12, 1519–1538, https://doi.org/10.5194/cp-12-1519-2016, 2016.

Haywood, A. M., Dowsett, H. J., Otto-Bliesner, B., Chandler, M. A., Dolan, A. M., Hill, D. J., Lunt, D. J., Robinson, M. M., Rosenbloom, N., Salzmann, U., and Sohl, L. E.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1), Geosci. Model Dev., 3, 227–242, https://doi.org/10.5194/gmd-3-227-2010, 2010.

Hill, D., Csank, A., Dolan, A., and Lunt, D.: Pliocene climate variability: Northern Annular Mode in models and tree-ring data, Palaeogeography, Palaeoclimatology, Palaeoecology, 309, 118–127, https://doi.org/10.1016/j.palaeo.2011.04.003, special Issue: Climate and seasonality in a Pliocene warm world, 2011

Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J. and Hill, D.J. (2008), A new global biome reconstruction and data-model comparison for the Middle Pliocene. Global Ecology and Biogeography, 17: 432-447. https://doi.org/10.1111/j.1466-8238.2008.00381.x

Oldeman, A. M., Baatsen, M. L. J., von der Heydt, A. S., van Delden, A. J., and Dijkstra, H. A.: Mid-Pliocene not analogous to high-CO2 climate when considering Northern Hemisphere winter variability, Weather and Climate Dynamics, 5, 395–417, https://doi.org/10.5194/wcd-5-395-2024, 2024.

Otto-Bliesner, B. L., A. Jahn, R. Feng, E. C. Brady, A. Hu, and M. Löfverström (2017), Amplified North Atlantic warming in the late Pliocene by changes in Arctic gateways, Geophys. Res. Lett., 44, 957-964, doi:10.1002/2016GL071805