Author response to referee 1 of EGUSPHERE-2025-5660

The Authors would like to thank Michiel Baatsen for taking the time to read and provide feedback on the manuscript, these comments will greatly improve the quality of the manuscript. Below is the response to the comments and details how they are addressed in the manuscript.
The referee comments are shown in black, with the authors response written in purple.

General comments:

Introduction: the paper needs a clearer statement on what is considered as 'the jet stream', from the context, I assume this is the polar/eddy-driven jet near the tropopause. In monthly mean zonal averages of zonal winds, the subtropical jet is usually much more prominent compared to the eddy-driven jet owing to the large difference in meridional variation. To clearly distinguish the eddy-driven jet in model output, one would either have to use sub-daily frequency or monthly means of the eddy fields (i.e. time-mean of U*U and V*V).

The revised manuscript contains a clearer statement on what is being used to define the jet stream and its variability. We used the maximum zonally averaged zonal wind speed which will remove some of the zonal variation within the monthly scale and will note this in the revised manuscript. The use of daily data is currently difficult given the lack of daily data from PlioMIP2 but may be possible for future studies to build upon this work by examining daily data from the PlioMIP3 ensemble.

In addition, there is no clear consensus in literature on the height/pressure level to study the jet stream, Many studies use e.g. 850 or 500 hPa levels, or the vertical maximum within any possible range between 900-100hPa. A clear example of this is the Abell et al. study mentioned on L94, in which the dust proxy is a clear indicator of low level westerlies (surface to ~850hPa), but much less intuitive regarding the tropopause jet stream. Without further clarity on these possibilities, it is difficult to adequately compare the conclusions made between different studies.

The differences in pressure levels used between studies has been noted as a limitation in the comparison between studies. In this manuscript we focus on 200, 500 and 850 hPa as these are the pressure levels common to all the models studied within the PlioMIP2 ensemble, and this has been clarified in the methods section of the revised manuscript.

Focussing on the Pacific and Atlantic basins does improve the ability to detect jet stream maxima in a zonal average sense. This is shown by the double het maximum over the Atlantic. As shown in Oldeman 2024, this double jet max can be related to a pattern of persistent anticyclonic wave braking over the North Atlantic Ocean (also shown in figure 3 of this manuscript). This is a known phenomenon in PI/PD conditions, which should be mentioned up front, as it is relevant to interpret shifts in jet latitude and strength.

The relationship between the variability in the jet latitude and wave breaking has now been included in the manuscript, both in the introduction and in the analysis of the variability. This has referenced Oldeman et al. (2024) and other studies also looking at Rossby wave breaking and its impact on double maxima seen in the North Atlantic and North Pacific regions. We consider how this relates to the data shown and the implications for the conclusions made.

From the introduction/methods section it is a bit unclear to me what the focus of this paper is. There is mentioning of earlier studies focusing on a single model, rightfully stating this as a main limitation. Further down, most of the focus in the methods is on new single-model experiments which seems a bit inconsistent. While part of the results are still on the ensemble, which are then

complemented with single model experiments, I feel that this is not stated clearly enough early on. My suggestion is not to change the overall setup, but to slightly alter the focus and/or the related communication.

A statement has been added to the end of the introduction in the revised manuscript to clarity the motivation for looking at single model analysis for some sections of this study. The main justification is the use of the forcing factorisation experiments to understand the drivers of change, and these experiments were only completed by a handful of models. Here we present new runs from HadCM3 but also make reference to the experiments from CCSM4-UoT and COSMOS in the relevant sections. The final paragraph of the introduction has been reformulated to the text below to improve clarity of the motivation. The revised manuscript will also contain more justification of the use of a single model at the beginning of Section 3.2. The revised paragraph in the introduction now reads:

"This study aims to examine Late Pliocene jet stream change, in the mean state and monthly variability, to understand more about the climate system from a multi-model perspective. We also present new HadCM3 simulations with an aim to understand the drivers of the change. This will provide a perspective on the usefulness of the Late Pliocene as a past analogue for future jet stream variability. Details of the simulations and analysis techniques used are in Section 2. Section 3 contains the results and an interpretation of the mean state jet, the variability in the jet stream and explores the roles of different forcing to the changes observed. In Section 4, we then provide a perspective on how future studies can build upon this work, enabling a deeper understanding on Late Pliocene climate, and how that can help inform future changes in the climate system."

In the methods section, it is stated that the analysis follows that of Li et al 2015 who consider winds at 850, 500 and 250hPa (edit: this is specified in the results section, maybe mention in the methods?). Please clarify what is studied here. In addition, taking a single maximum from the zonal average leaves out all of the zonal variability in jet latitude and strength within a single time frame. This is a choice that may be justified, but please be more clear and motivate why. As shown in Oldeman et al. 2024, double jet maxima may considerably complicate the analysis and interpretation of jet strength. In addition, there is no clear statement on which time means are considered. The study considers monthly data, but are these averaged over the winter season (DJF) for each year? For variability, there is a brief mentioning of January means, are these then different from the time-mean analyses?

The pressure levels considered (200, 500 and 850 hPa) have now been clarified in the methods section.
The revised manuscript now contains a clearer justification on using a single maximum (" We use monthly, zonally averaged data which smooths out some of the higher frequency features making it sufficient for mean state analysis") and references the impact of double jet maximum. In particular the double jet maximum is explained with a focus on the North Atlantic and provides further rationale as to why the variability analysis in our paper is focused on the North Pacific. The mean state analysis is focused on Northern Hemisphere Winter (DJF) in order to increase the amount of data going into the analysis. For the analysis of variability, only January is used to remove some smoothing and to more accurately see changes in the speed and latitudinal position of the maximum zonal wind speed. The position of the maximum zonal wind speed may be more relevant to changes in mid-latitude extremes. The periods used have been clarified in the text and in the figure captions.

Subsection 3.2 in the results needs a lot more care regarding the 3D structure of subtropical and eddy-driven jet streams, as well as how this temporally varying structure is represented in the time means at a single pressure level. The link between temperature gradients and wind magnitude is implied but mostly lacks a proper explanation or background. These results are still relevant and the analyses seem sound, but considerably more care should go into motivating as well as interpreting them.

Section 3.2 now contains a more detailed analysis of the vertical studies of both the zonal winds and temperature. Figure 5 has been replaced with Figure R1 showing the forcing factorization of the vertical temperature profiles. And Figure R2 has been added to show the vertical structure of the zonal winds. Section 3.2 has been revised to discuss these vertical profiles and the link between temperature and wind speeds. We also find a change in the geopotential height over the Aleutian low region (Figure R3) and the revised manuscript notes the similarities between this study and that of Menemenlis et al. (2021) and Oldeman et al (2024). Figure R3 has been included in the revised supplementary material.

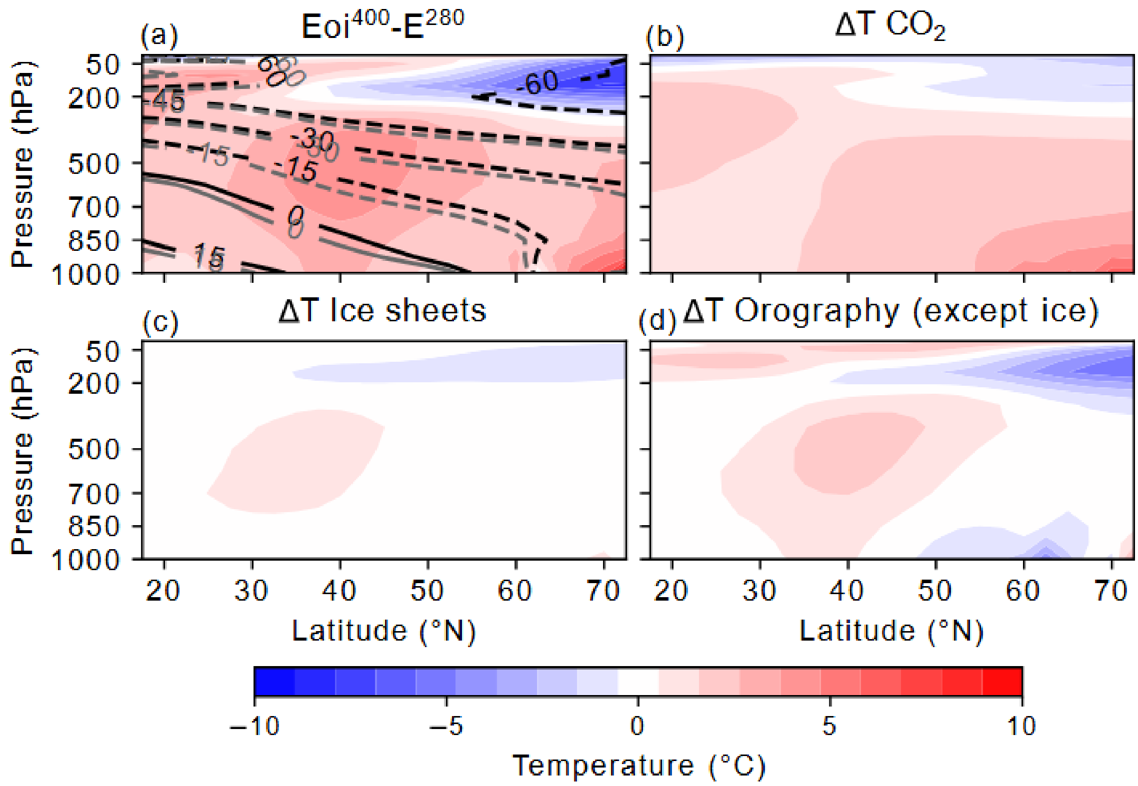

Figure R1: Change in the DJF North Pacific Temperature Profiles in HadCM3 and the contribution from CO2, ice sheets and orography. The grey contours show the pre-industrial (E280) values and the black contours show the Late Pliocene (Eoi400) values.

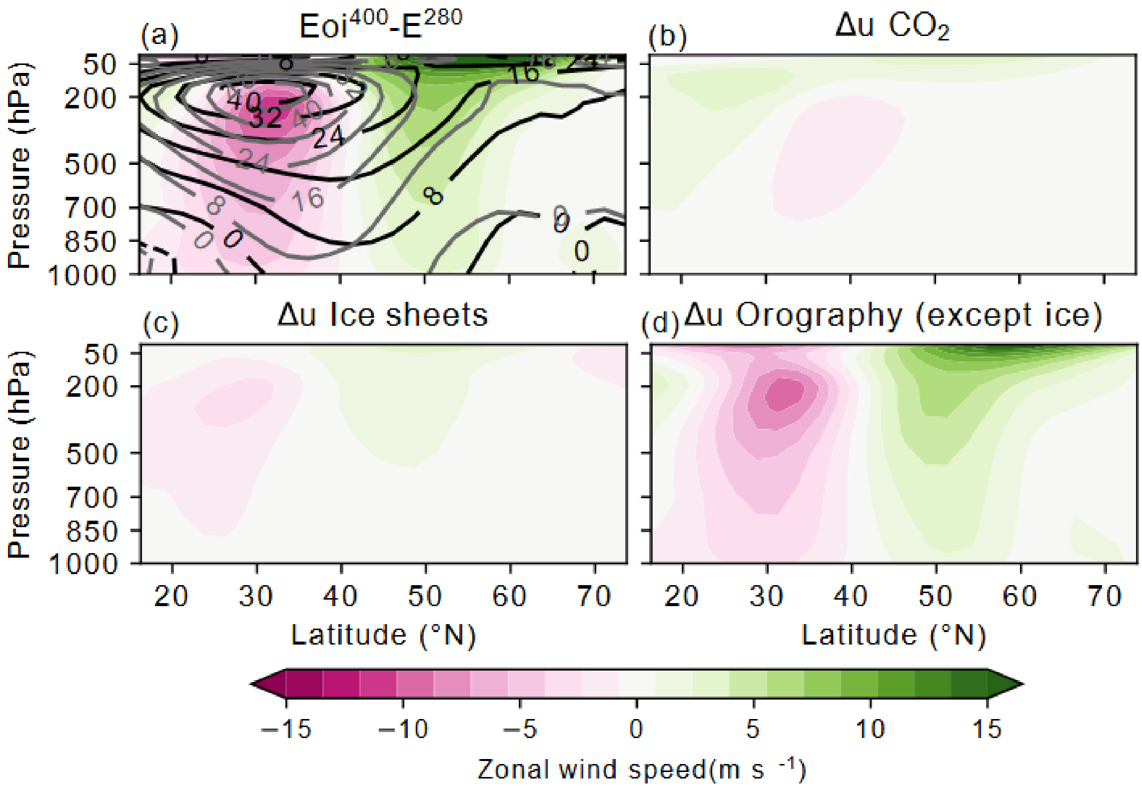

Figure R2 Change in the DJF North Pacific zonally averaged, zonal winds in HadCM3 and the contribution from CO2, ice sheets and orography. The grey contours show the pre-industrial (E280) values and the black contours show the Late Pliocene (Eoi400) values.

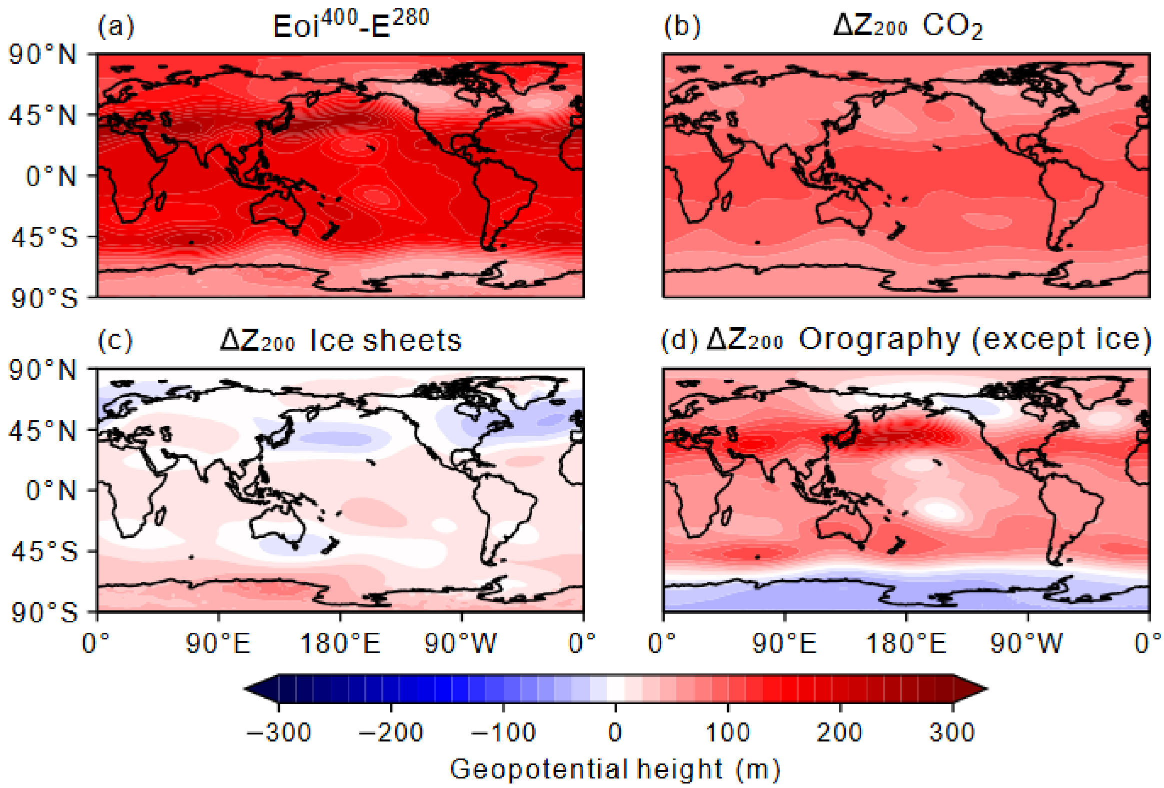

Figure R3 Change in the DJF Geopotential height in HadCM3 and the contribution from CO2, ice sheets and orography.

We have looked at the 3D structure of the zonal winds in HadCM3 and examined how the vertical structure may change the results of the variability analysis. Over the North Pacific regions, there is indeed vertical displacement of the time-averaged zonal wind speed (Figure R4). Examining the variability of this through time there are more occurrences of the maximum zonal windspeed occurring the 250hPa level opposed to 200 hPa (Figure R4) which is analysed in the manuscript. However, there are differences between models on what pressure levels the data is provided on with 200, 500 and 850 hPa being common to all models which is one reason that these levels were chosen. In order to include as many models as possible into the variability analysis we continue to use 200 hPa in the revised manuscript but note it as a specific source of uncertainty in the results presented. Figure R4 are included in the revised supplementary material to show these changes

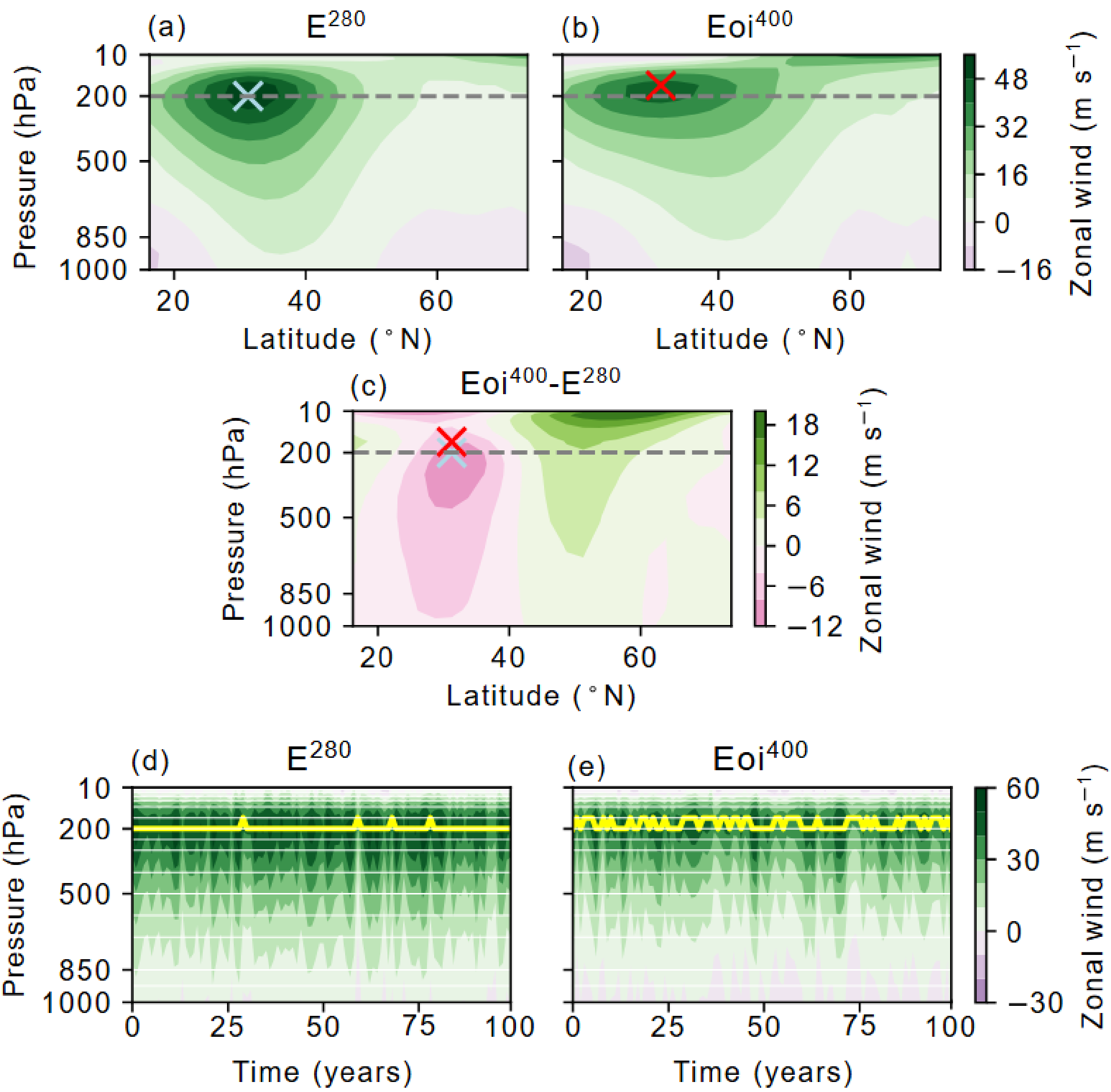

Figure R4: (a-c) HadCM3 North Pacific zonal wind speeds for Northern Hemisphere winter. The results are presented for the pre-industrial control (E280), the Late Pliocene simulation (Eoi400) and the difference between them. The dashed grey line represents the 200 hPa layer and the blue and red crosses mark the location of the maximum zonal wind speed for E280 and Eoi400 respectively. (d-e)HadCM3 North Pacific zonal wind speeds in the North Pacific (averaged latitudinally between 20N to 40N) for 100 Januarys of the pre-industrial control (E280) and Late Pliocene simulation (Eoi400). The yellow line shows the location of the maximum zonal wind speed.

Subsection 3.3.1 needs some better structure; it quickly jumps between a large number of figures and generally lacks a solid motivation, interpretation and above all connection between the different results mentioned. This makes it tough to see the general picture and main message here.

The motivation in subsection 3.3.1 has been more clearly worded: *"Using the new forcing factorisation experiments the contribution of each boundary condition change can be assessed to understand if the drivers of the mean state also drive changes to the variability."*
The main message in this section is to link the change in the variability of position of the maximum zonal winds to a change in the boundary conditions. We find that the Late Pliocene orography (including land sea mask and vegetation changes) is a main contributor to this change in the zonal winds. An analysis of sea ice is included as in past literature sea ice has been linked to mid-latitudes

circulation. Furthermore, the results showing the grouping of the experiments by which orography is used is interesting as it also provides a possible explanation of the changes seen.

Specific comments

L11 'This is important as …'Could the statement be further clarified or specified by reworking the sentence to make it a bit less vague?

This statement has been rewritten to improve clarity: *"This change in jet stream variability in the Late Pliocene could lead to a change in the distribution of temperature and precipitation which could have implications for how proxy data and model simulations are compared."*

L58 At upper levels, there is indeed an enhanced meridional temperature gradient in a warmer world (particularly through greenhouse gas-induced warming), but this mainly results from lower stratospheric cooling at higher latitudes as opposed to upper tropospheric warming at lower latitudes. Both can be at the same height/pressure level, owing to the meridional structure of the tropopause. Please clarify this.

This has been clarified in the revised manuscript

L113 CESM2 and CCSM4_Utr are not used due to a difference in coordinate system; please explain? The latter was used to study the jet stream in Oldeman et al. 2024, so apart from being possible, including this into the study would help compare the results.

We agree that it would have been interesting to include these models in our analysis, however we had difficulties in processing the data on the irregular vertical grids, and chose to not include them. Some of the other PlioMIP models were also not included due to difficulties in processing the data.

The 11 models that we have used provides a reasonable sized ensemble for analysis and a good indication of inter-model variability.

L115 Using the PlioMIP3 nomenclature does not seem intuitive to me, as this study considers PliMIP2 model output. Would it make sense to list the PlioMIP3 nomenclature instead and use the PlioMIP2 one in this paper to make it more comparable to previous work?

We agree that the PlioMIP2 nomenclature is more appropriate for this study. The revised manuscript will use the PlioMIP2 nomenclature. We still refer to the time period as being the 'Late Pliocene' but chose to not abbreviate it so there is no confusion with the LP PlioMIP3 experiment.

L128 please specify 'a good climate'

The revised manuscript has been updated to avoid the term "good climate". The text now reads: "*It simulates a Late Pliocene warming of 2.9°C, close to the MMM (3.2°C), within the model range (5.2-1.7°C) and is in reasonable agreement with proxy reconstructions (Haywood et al., 2020)."*

L130 I was puzzled by the statement on the vertical levels for a moment, until I noticed I missed the 'un' in unevenly spaced levels. Maybe rephrase slightly for clarity?

This line has been revised in the manuscript to "and 20 unevenly spaced vertical layers and a time step of one hour" to improve clarity.

L158 ERA5 consists of a single model (i.e. IFS), rather than models? In addition, the 85 years spanned by ERA5 could be considered as similar to the 100 years in the PlioMIP ensemble? Using different reanalysis datasets is always helpful for a more complete comparison, but it is at least as

important to consider its reliability (in addition to reduced observation methods/counts) in the pre-WWII period.

The change from models to model has been implemented in the revised manuscript. For this analysis we use 1940-2015 for ERA5 and 1916-2015 for NOAA-CR20 and have included this this in the revised manuscript. We have also added in a statement referring to the reliability of reanalysis data representing older time windows: "*The reanalysis data will also contain influences of anthropogenic climate change, so is not a perfect comparison to model runs in climate equilibrium, however reanalysis data becomes less reliable the further back in time due to limited data sources so the more recent reanalysis data is used*."

L170 I completely miss a statement on which vertical level or pressure level is considered to determine the jet stream.

We manuscript has been revised to clarify the pressure levels examined by adding the following text: "As in Li et al. (2015), the zonal wind is examined on three pressure levels (850, 500 and 200 hPa)"

L196 I am uncomfortable with the use of subtropical/polar jet here. As shown in figure 3, the jet stream pattern over the North Atlantic can be linked to persistent anticyclonic wave breaking. This causes the eddy-driven component of the jet to be dominant over the western/central part of the basin and the 'conventional' subtropical jet to regain strength over the eastern part. As you are considering time means of zonal wind at 200hPa, the analysis is strongly biased towards showing the subtropical jet. The pattern over the Atlantic is an exception to this rule that deserves much more care and attention interpreting the results.

The use of subtropical and polar jet has been removed and replaced with a discussion around Rossby wave breaking.

L208 Linking the 850hPa and 200hPa levels is indeed useful for proxy comparisons, but doing this based on a qualitative comparison between both fields in the MMM is not very robust. A correlation between different models and/or years would provide a much stronger argument.

The revised manuscript now only references the changes to the zonal winds at 850 in relation to proxy data, however the vertical profiles of wind speed added to section 3.2 show the change in zonal wind speed is consistent through the depth of the troposphere in HadCM3.

L220 HadCM3 is noted as a clear outlier when looking at the reanalysis data, this strongly advocates some further discussion on the interpretation of the model-specific analyses further down.

The manuscript has been revised to contain a discussion about where HadCM3 sits in the ensemble in terms of change to zonal wind speeds as well as an expanded section comparing HadCM3 to the other models with the forcing factorisation experiments (CCSM4-UoT and COSMOS).

L235 Apart from the CO2 response being weaker, it is also opposite in sign compared to the other 2 forcings.

This has been noted in the revised manuscript: *"The change in wind speed due to CO2 is small (and opposite in sign to the other forcings), with the largest contribution being orography (including land-sea mask and vegetation) and ice sheets following as the second leading cause in the change."*

L236 I assume looking at the temperature responses is motivated by thermal wind balance, but I do not see this being mentioned? In that case, looking at the temperature response integrated

over the atmospheric layer below would be more suitable. Furthermore, comparing the meridional temperature gradient response to CO2 and ice sheets versus the wind speed response, seems to be rather inconsistent. This discrepancy (if correct) in the results seems to be missed out on or ignored altogether.

Section 3.2 in the manuscript has been revised to contain a more in-depth look into the temperature changes and vertical profiles of the model results, and how they impact the zonal winds including an analysis of geopotential height change.

L237 Please be more specific regarding 'upper polar/upper tropical'regions, as this may imply anything from the top of the boundary layer to the top of the atmosphere.

Due to the change made in section 3.2 to contain the vertical profiles, this statement is no longer included in the manuscript.

L240 A weaker temperature gradient may lead to a weaker and poleward jet stream; what is the latter statement based on? Does this hold for the jet near the tropopause, or is it only valid at 850hPa? The first argument seems to be the complete opposite of what is shown in the figures, with generally enhanced temperature gradients at upper levels. Furthermore, over the Atlantic Ocean, there is no clear change in strength or average latitude, as the breaking wave pattern is reduced in strength and converges towards a single jet latitude in LP compared to PI. In addition to a more general shift, the opposite is seen over the North Pacific, both of these responses are consistent with Oldeman 2024.

As section 3.2 has been changes to include a discussion about the vertical profiles, this statement is no longer included in the revised manuscript. This figure is no longer include in the manuscript and had been replaced with a vertical temperature profile over the North Pacific.

L245 Also cite the Otto-Bliesner 2017 paper here?

This citation has been added to the revised manuscript.

L257 This sentence is rather tough to understand; are you talking about model differences in general, or specifically between CCSM4 and HADCM3? Why would a difference in the mean automatically imply the same for variability? (I'm not saying it does not, I am just unsure why).

Here we are referencing the range in response to Late Pliocene forcings across the PlioMIP2 ensemble. We have clarified that due to the range in model response to the mean state, there may also be a range in the temporal variability across the ensemble which is why we have chosen to perform a multi model analysis of the variability.
There are less models with the data available to examine the causes of the changes in the temporal variability so we present HadCM3 in the main paper and the other two models (COSMOS and CCSM4-UoT) in the supplementary material.
This section now reads *"Although jet stream variability has been investigated in similar manner before within CCSM4-Utrecht only (Oldeman et al., 2024), as discussed in section 3.2, there are large differences in the mean state shift across the ensemble. This range in response in the mean state may also suggest differences in simulated variability across the ensemble, highlighting the importance to examine variability in a multi-model setting."*

L259 Does this consider the full NH, as opposed to Atlantic/Pacifc before? What would be the reason to differ from the previous analyses? Considering the large differences between the basins

certainly limits the interpretation of these results. Edit: I see the figure caption mentions North Pacific, please clarify in the text?

This section does indeed refer to the North Pacific region. The subheading for section 3.3 has been renamed to 'Jet stream variability in the North Pacific' to improve clarification. We chose to focus on examining the variability in the position of the maximum zonal wind speeds over the two regions due to the differences seen in the mean state between the two regions. Due to the double peak occurring in some model runs over the North Atlantic, which complicated the analysis and interpretation, we choose to focus on the North Pacific for this study.

L276 There needs to be a clear explanation of what is considered as 'jet stream' variability, as the link between strength and variability is made multiple times in this work (and shown to be significant in Figure 9). There is, however, a substantial difference between spatial variations (i.e. 'wavy' jets) and the temporal variation of the position of the maximum in zonal average zonal wind speed. Both may be related, but this would need some proper motivation.

Here 'jet stream variability' relates to the variability (standard deviation) in the position of the maximum zonally averages zonal wind speed. In section 2.2-Analysis of model simulations, we have edited the sentence to read "To measure the variability of the jet stream latitude (referred to as jet stream variability), we calculate the standard deviation of the latitude of maximum wind speed in the month of January.". Furthermore, we have clarified that the variability related to the variability in the position in this line : *"In HadCM3, the North Pacific Late Pliocene simulation jet stream is weaker and more variable in position than in the pre-industrial simulation (Fig 8) in agreement with the majority (7 of 11) of the models."*

L292 I have seen the suggestion of linking a weaker meridional temperature gradient to a weaker and more wavy jet stream before, but there does not seem to be a clear physical mechanism nor observational evidence for this? Note that, regarding the above statement, wavy jets are not the same as temporally changing latitudes of the zonal wind max. Making any claims on wavy jets would require a much more detailed analysis of spatial patterns at high temporal frequencies and/or eddy components of velocities and fluxes.

We have edited this sentence to remove the reference to the wavy jet and have added more references regarding the link between meridional temperature gradients and weaker zonal wind speeds: *"This Arctic amplification, leading to a weakening in the meridional temperature gradient, could create a slower jet stream (Francis and Vavrus, 2015; Smith et al., 2022)."*

Figures:

Figures In general, please be more consistent with the sizes of e.g. fonts and colourbars between the different figures. Also: add lat,lon coordinates to the spatial figures?

The figures have been edited to have a consistent font size and to contain lat/lon coordinates.

Figure 1 Please add a vertical dashed line or grid line showing the zero value to interpret that LP-PI change. Also consider scaling the change (e.g. x10) for readability. Minor suggestion: while I appreciate the consistent scaling, the range in wind speeds can be reduced considerably for the 500/850hPa panels, improving readability. If consistent scaling is desired, you may adjust the panel width accordingly as well. Is the figure showing DJF, January, or annual mean?

These changes have been implemented with the exception of the scaling of the difference line as this did not improve clarity of the figure. This figure is showing DJF and that has been clarifies in the figure caption

Figure 3 Please make use of a diverging colourmap, or a shift in colour for values that are below zero i.e. showing easterly winds. Again, also indicate whether this is showing boreal winter, winter in general, or something else? The scaling of the difference plots could also be reduced to improve clarity?

These changes have been implemented as suggested

Figure 4 Adding just a single contour for the PI reference (e.g. 30m/s) would really help interpret whether the changes mean a change in strength or a spatial shift in these panels.

These changes have been implemented as suggested

Figure 6 The colourbar in combination with the contour lines is pretty rough. In addition, the relevance of this figure in the main text seems limited, as this is only used to argue that the AMOC is indeed stronger in the LP versus PI experiments, being consistent with previous work? The full 2D structure of the overturning stream function is of limited added value to this study.

This figure has been moved the supplementary. The contours are retained to allow for the changes between the subplots to be seen more clearly.

Figure 8 While this is a rather intense multi-panel figure, I do appreciate the complete overview among models. For comparison, it could be helpful to add the numbers of mean and variability for each case, which are otherwise not shown?

The mean latitude and the variability have been added to each subplot as suggested.

Errors/typos

L28 Northern Hemisphere? (also on L244)

L30 was simulated?

L73 not one?

L92 LP stands for Late Pliocene?

L168 by-linearly?

L190 use \citep?

L208 in comparison to the change?

L217 slightly akward sentence, maybe rephrase?

L226 a possible causes, jets stream

L270 redundant period? Fig8 caption for in January

L283 CO2-driven?

L284 to c achange … could be application to

L290 feedbacks positively

L326 Oldeman 2024

The authors thank the reviewer for highlighting these errors and they will be corrected in the revised manuscript.

Author response to referee 2 of EGUSPHERE-2025-5660

The Authors would like to thank Anonymous Referee #2 for their time to read and provide feedback on the manuscript, these comments will greatly improve the quality of the manuscript. Below is the response to the comments and details how they are addressed in the manuscript.

Referee comments are shown in black, and the authors response are shown in purple text.

Overarching comments:

1. Relevance of the research and link to proxy data

From my personal point of view the discord between mid-Pliocene jet stream position being more northward in an environment with reduced meridional temperature gradient, and understanding based on modern observations suggesting an equatorward shift in the jet stream (line 55ff of the manuscript) is relevant. Could the authors shed more light on this problem based on the simulations that they analyse? Furthermore, the manuscript refers at many locations the link to proxy records, illustrating that combination of proxy records and models could improve our understanding of past extreme events. Nevertheless, the major take home message that I gain from these statements are problems with recording / reproducing extremes in both methods. In my opinion this proxy-model topic stays in general rather unclear. More examples would be helpful to better understand the relevant points that the authors would like to make here. I provide some example below.

The poleward shift in the Pliocene jet stream in the North Pacific is due to changes in orographic boundary conditions creating a change in temperature patterns. The revised manuscript contains a more detailed analysis of the vertical profile of the jet. This highlights the small role CO_2 plays in the change of the westerly winds making it hard to draw more comparisons to modern/future climate. A reference to a study on modern climate (Woolings et al., 2023) has been added to section 4 to clarify the similarity between the change and the mechanism. It would be difficult to comment more on this change without directly comparing to future simulations.
Proxy records are mentioned regarding their use in evaluating the mean state response, which is one motivation for examining the climate of the Late Pliocene as there are many proxy records. As there are no proxy records of extremes for this time period, we rely on modelling efforts to understand the variability and extremes. So, the motivation is to start by looking at large scale features of the climate system (like the jet stream), and how they may be represented in models and how it compares to proxy data. Here we find that the mean state is a weaker jet stream which may be related to a more variable jet stream (Figure 9). From this work it may then be possible to comment on how this variability translates into climate extremes in terms of precipitation and temperature distributions although to fully capture these changes, daily data may be required which should be provided as part of the PlioMIP3 protocol. This motivation has been clarified in the revised manuscript.

2. Consideration of jet stream variability across two different regions (North Pacific and North Atlantic)

The authors separate the analysis across two geographic regions, the North Pacific and the North Atlantic region. In fact, when studying the figures, it is not always clear which region is shown in analyses. This should be clarified. Furthermore, at least to me it remained a bit unclear from the manuscript what is gained by studying both different regions, and to which extent the findings of each region have different relevance for our understanding of climate dynamics. The respective

text in lines 170ff remains a bit vague with regard to this motivation. Since the authors focus a lot on AMOC, and AMOC having its major impact on the heat distribution in the North Atlantic region, I could imagine that the North Atlantic region's jet stream may be more relevant for various dynamics in Europe, for example. Nevertheless, I get the impression that more analyses are presented for the North Pacific region (although, as noted above, I might be wrong as it is not always clear from figure captions which analysis region is actually shown). Please consider during your revisions to make the motivation for focusing on each of these two regions clearer, also highlighting the specific relevance that findings derived for each region's jet stream have for our understanding of the (Pliocene) climate system.

We have revised the manuscript to be clearer in the figure captions and in the text what region is being studied. Section 3.3 has been renamed to 'Jet stream variability in the North Pacific' to improve clarity about what is being discussed in this section. In the mean state analysis, we present the two different regions as they have two different patterns of zonal wind speed, and their separation makes it easier see the jet stream maximum. Due to the presence of a double maximum over the North Atlantic region in some models, which complicates the analysis and model comparison of the variability, we focus on the North Pacific. This has been clarified in the methods section of the revised manuscript. The inclusion of the AMOC is to signal one possible mechanism for more heat transport towards the Northern Latitudes along with changes to sea ice, and vegetation patterns.

3. Terminology

I am not sure whether readers outside the PlioMIP modelling community find it easy to link simulation names to specific time periods or to the employed forcing factorization. Consequently, I suggest to the authors to reconsider the chosen terminology along the following lines:

The choice was made to use the PlioMIP3 terminology for this manuscript so that it would align with future papers published using PlioMIP3 data. However, we appreciate the possible confusion the reviewer highlights. We are aware that there is also possible confusion when referring to LP – the late Pliocene, and LP – the model experiment. To clear up this confusion we address the following comments.

3.1 Definition of the main time period Late Pliocene: The authors chose the term Late Pliocene that comes with the update from PlioMIP2 to PlioMIP3 and define this in their lines 20-21. For clarity to the reader, here it would make sense to also refer to Haywood et al. (2024) regarding the renaming.

We retain the use of the term 'Late Pliocene' when referring to the time period of study, this aligns with the PlioMIP3 name but also reflects other studies done on this time period. We have now changed the text to "*One time period well suited to this is the Late Pliocene, specifically Marine Isotope Stage KM5c, 3.205 Ma (within the mid-Piacenzian Warm Period) (Haywood et al., 2024)".* In the original manuscript, we did not abbreviate this to LP to retain separation between the formal geological time definition, and the model experiment, we chose to keep this unabbreviated so that there is not future confusion as the LP (the experiment name) becomes more common in the literature.

3.2: While the authors provide Table 2 as a service to the readers towards linking PlioMIP2 and PlioMIP3 terminology, my feeling is that non-experts of PlioMIP terminology could find comprehending section 2.1.1 a bit challenging; in particular, you describe results from PlioMIP2 simulations and refer to them with the newer PlioMIP3 term Late Pliocene. I think that the link to

simulation Eoi400 (and its overlap with simulation LP) should be made more clear here. Towards improved clarity, I suggest to move text currently written in lines 115-117 to the beginning of the section.

In the revised manuscript, we now use the PlioMIP2 nomenclature but retain Table 2 to allow for comparisons of future studies from PlioMIP3 as the core experiments between the two phases are the same there is no difference in the Eoi400 and LP experiments so this mean state analysis may be relevant for future studies with PlioMIP3 data. Due to the change in the nomenclature, line 115-117 is no longer required in the manuscript.

3.3: I note that the link of boundary conditions to PlioMIP3 names is not always obvious from the PlioMIP3 terminology. For example, in simulation PI_lp-orog the term PI identifies PI icesheets, while in simulation PI_lp-ice it refers to PI orography (your Table 2). I think it may be helpful if you provided in your forcing factorization analysis clear indications as to the affiliation of all relevant boundary conditions to PI or LP background states. For example, in Figure 5 you could add the text "(except ice sheets)" to the heading of panel d. Comprehension of Figures 6, 10, and 11 could likewise profit from illustrating which parts of the boundary conditions are LP vs. PI by adding respective information in brackets after simulation names. This would help in linking results to the state of specific boundary conditions.

We have changed to the PlioMIP2 terminology which should improve the clarity of the text and figures. We have also added in '(except ice)' into the subheading of figure 4d and 5d.

3.4: Terminology of the term "orography" in the context of boundary condition changes. In my personal experience the term orography is most often used in referring to elevations (bathymetry or land height surface). In PlioMIP terminology, on the other hand, the term refers more generally to combined changes of elevations, land sea mask, and other characteristics outside of ice sheet regions (maybe one could call this combination of changes called "orography" as modifications of geography outside ice sheet regions). Since in the context of this study the term orography is often used by you to implicitly refer to differences in land-sea mask that influence the expression of the AMOC (e.g. line 243 ff), I think it may make sense to explicitly spell out those characteristics of orogography that are of relevance in a specific context. In line 245 it would be particularly important, towards clarity, to replace the term "orography changes" by "gateway changes", and in line 266 by "land-sea mask changes". Also in lines 279ff it could be considered which occurrences of the term orography rather refer to land-sea mask changes specifically.

We have added the following text into the methods section to clarify that when we use the term 'orography' we are referring to orography, land-sea mask and vegetation. *"Following the nomenclature of PlioMIP2 (Haywood et al., 2016), 'o' represents Late Pliocene orography (including land-sea mask and vegetation, hereafter referred to as orography) outside of ice sheet region"*

Furthermore, we have also clarified at other points in the text that the change in the land-sea mask is responsible for the observed changes.

"The experiments employing Late Pliocene orography (including land-sea mask changes) show a stronger AMOC than the experiments with pre-industrial orography."

"Since we expect the differences observed in the jet to be partially caused by land-sea mask changes, we do not expect a large change in the jet stream in these models, as is seen in Fig"

Relating to lines 279ff we keep the term orography as here we also mean to refer to the vegetation changes also impacting the temperature gradient.

3.5: interannual variability vs. variability vs. differences. I note that I sometimes got lost in the descriptions of variability of jet stream vs. differences in the variability over time and between different models. Please carefully review text passages where you describe and discuss related findings and make sure that the meaning is always clear. Example: You speak of interannual variability heavily in the abstract and once in the outlook (line 334), but that term actually never appears in any other part of the main manuscript. What is the difference between interannual variability addressed in outlook and abstract vs. variability described in other sections? Another example, line 293-294: Here you speak of variability in the jet stream across the 8 simulations for these models. Do you never here to interannual variability of the jet stream in each simulation, or to differences between jet stream expression between different simulations? Also see other locations.

In the manuscript, the term variability refers to the interannual variability in the position of the maximum zonal wind speed. In the revised manuscript, care has been taken to ensure that the difference between variability in the jet latitude, and variability across models is clear.

3.6: The authors focus on boreal winter when analyzing the jet stream. Yet, it is not always clear from figure captions which seasons the results present (see specific comments below). Furthermore, sometimes the analysis uses DJF means, sometimes just velocities of January. Please provide reasoning for this choice in the revised manuscript.

For changes in the mean state, we look at DJF to gain as complete a picture of the mean state as possible. When looking at the variability, we choose to look at one month (January) so that we don't average out differences in an unreasonable way. Including more than one month per year may introduce a bias if there is a relationship between the jet metrics between these months ((i.e. the jet in January could be correlated to the jet in February). The choice of using month to study the variability in consistent with previous studies (Oldeman et al. 2024). The following text is added into the methods section *"Only one month was studied for the variability metrics to not smooth any changes in latitude across a season.".*

3.7: the terms "mean" vs. "multi-model mean (MMM)" seem not to be used consistently. For example, line 200 ff. I think one appearance of "mean" rather refers to the MMM than to the mean state of the jet stream, which is also a topic of interest addressed in the study.

The manuscript has been updated to improved consistency in the use of multi-model mean (MMM). We have retained the use of 'mean state' to refer to the multi model mean, mean state.

4. Abbreviations

Abbreviations are not always consistently defined and used. Late Pliocene is sometimes abbreviated, sometimes it is not. Please use a consistent strategy across the manuscript and supplement. In figures, please make sure that abbreviations in figure elements are always defined in the figure caption - this is relevant in particular also for the supplement, where selected figures may be studied more outside the context of manuscript-defined terminology.

As described in response to comment 3.1 we have now removed all appearances of ‘LP’ in reference to the Late Pliocene, but retain it in table 2 (Table 1 in the revised manuscript) as it refers to the experiment name.
The figures now contain definition sof the abbreviations used in the figures.

5. Clarity of formulation, argumentation

Below I will highlight some text passages where the reasoning of the authors stayed unclear to me or where text elements seem to me disruptive to the overall tread of argumentation of the manuscript. Please reconsider your argumentation aims in such cases and clarify text accordingly.

Thank you for taking the time and care to highlight these sections. We have addressed these in the revised manuscript, details on these change follow each specific comment in the text below.

6. selection of indices in the comparison of large-scale climate metrics to jet stream changes and variability

In line 226ff and Fig. S5 authors analyze the relationship between change in jet stream latitude vs. ESS/ECS. The authors do not find a strong relationship. To me the motivation to choose ESS/ECS as a metric did not quite become clear. Maybe the authors could clarify their reasoning in the text, and generally expand the related text section. Furthermore, ESS and ECS describe the response of >global< climate to changes in forcing and boundary conditions. Yet, in their motivation and also in their analysis, the authors highlight the relevance of the meridional temperature gradient for the expression of the jet stream. Taking this fact into account, wouldn't a comparison between jet stream and polar amplification of a model appear more promising when aiming to understand the link of jet stream changes to large-scale climate metrics of climate models? A stronger polar amplification will relate to a larger reduction of the meridional temperature gradient, and therefore one could also expect to find a clearer relationship between jet stream and polar amplification.

As the ESS/ECS ratio describes the impact of CO2 vs other boundary conditions on global temperature. The forcing factorisation showed that non-CO_2 boundary condition were the most important for the zonal wind speed changes. We use ESS/ECS to assess whether models that show a relatively stronger effect of non-CO2 forcings on the Pliocene climate also show a large change in the zonal winds.

The revised manuscript contains a figure showing the vertical temperature profile over the North Pacific region which demonstrate the relationship between temperature and zonal wind speed. The relationship between the zonal wind and polar amplification has been examined but no strong or significant correlation was found so it not included in the revised manuscript.

7. typography

Across the manuscript spacing between scalars and physical units should be observed in both text and figure elements (pressure levels are consistently referred to without spacing).

Care has been taken to ensure improved typography in the revised manuscript and the figures have been updated with the correct spacing between the values and the units.

Specific comments:

line 11-13: "This is important as these differences might suggest a shift in the distributions of climate variables, such as temperature and precipitation, which could have implications for how proxy data and model simulations are compared." Towards increased clarity, please be more specific regarding the reference states that you compare here with the term "shift". Do you mean Pliocene vs. future or a different quality?

This has now been revised to *"This change in jet stream variability in the Late Pliocene could lead to a change in the distribution of temperature and precipitation which could have implications for how proxy data and model simulations are compared."*

line 45-49: Here the reasoning is unclear to me. The authors refer to difficulties in deriving proxy records towards reconstructing extreme events. As a remedy, authors highlight the value of climate models towards increasing our understanding of drivers of extreme events and of their change to then motivate hypothesization on the impact of such events in the proxy record. Thereafter, the authors again highlight the value of studying a climate, where ample proxy data is available, towards understanding the skill of climate models in simulating extreme events. Yet, the authors just highlighted problems with reconstructing extreme events from the proxy record. How is this problem being overcome in the model-data comparison? I hypothesize that the authors would like to hint here at gaining a better understanding of changes in large-scale climate patterns and variability based on climate models towards subsequently enabling the study of the transfer of such changes into proxy archives via proxy system models. Yet, at least to me such an argumentation seems to remain unclear from the manuscript. Please clarify.
As climate extremes are difficult to see in the proxy record, we chose to approach this problem by starting with a broader view of the climate system and consider the mean state changes (which have more opportunity to be captured in proxy system) and then narrow to shorter temporal scales. The reviewer is correct in saying that understanding large scale climate features will help to improve understanding of variability and extremes. Even though we cannot currently validate extreme events with proxy data, the large scales features within the simulations can be validated. We can then use the simulations to assess the extremes that a climate with these large-scale features would experience.

We will add the following text to clarify this: "*Furthermore, looking at extreme events in simulations of a warm climate where the mean state has been validated against proxy data (McClymont et al., 2020; Tindall et al., 2022), can help us understand how climate models represent large scale climate features. Understanding the large scale features, and how well a model captures them, is a necessary first step to understanding the variability, extremes and their drivers seen in the climate system.*"

line 92: abbreviation LP undefined. It may be sensible to also here once more provide the link to the MPWP for readers not yet familiar with the PlioMIP3 terminology

We have changed this to be 'Late Pliocene', as we have reverted to the PlioMIP2experiemnt names, we do not add in a link to the PlioMIP3 terminology as we no longer use the PlioMIP3 experiments which should improve the clarity.

line 92-96: this reference of proxy-based findings to jet stream variability seems a bit out of place. One could shorten this to the first two sentences of that paragraph. The remainder of the text could be better linked to the remainder of the manuscript, or maybe just moved to a later section where the focus is again on proxy-based evidence (currently this is in lines 208ff, but I note that

also this text section seems a bit out of place in the results section and could be better placed in the discussion)

We have reworked this section of the introduction to remove the second half of the paragraph as suggested, and added in more general information about model comparisons to proxy data (line 96-103 in the revised manuscript)

We chose to keep the text in this location (now lines 242ff in the revised manuscript) of the results and interpretation section as the proxy data is better suited to the mean state analysis rather than the following sections on the variability. The new text now reads *"Abell et al. (2021) examine changes in the strength and position of the jet stream using dust flux records in the warmer Pliocene climate compared to cooler glacial climates. They report that in warmer climates, like the Late Pliocene, the jet stream is poleward and weaker compared to cooler climates matching what is found in this modelling study."*

line 102: Text of the introduction seems to end abruptly here. Also refer to the conclusions section, and maybe provide a final sentence as a transition to the methods section, maybe regarding the motivation of your study?

The final section of the introduction has been revised to more clearly state the motivation of the study and to provide a smother transition to the rest of the manuscript. The revised section now reads: "*This study aims to examine Late Pliocene jet stream change, in the mean state and monthly variability, to understand more about the climate system from a multi-model perspective. We also present new HadCM3 simulations with an aim to understand the drivers of the change. This will provide a perspective on the usefulness of the Late Pliocene as a past analogue for future jet stream variability. Details of the simulations and analysis techniques used are in Section 2. Section 3 contains the results and an interpretation of the mean state jet, the variability in the jet stream and explores the roles of different forcing to the changes observed. In Section 4, we then provide a perspective on how future studies can build upon this work, enabling a deeper understanding on Late Pliocene climate, and how that can help inform future changes in the climate system*."

line 111: Should also Hudson Bay and Western Antarctic land sea mask changes be explicitly mentioned here due to their importance for AMOC dynamics (refer to Otto-Bliesner et al. (2017) wrt. Hudson Bay)

The revised manuscript contains an expanded list of the land sea mask changes to include changes to the Hudson Bay and Antarctic land sea mask.

line 113-114: the argument that "CESM2 and CCSM4_Utrecht are not examined here due to a difference in coordinate systems" remains unclear. For clarity, please explain here what exactly differs so that those models' simulations are ignored in your study.

These two models are not included in the mean state analysis as the vertical grid that the model data was provided on created a challenge when attempting to re-grid this to specific pressure levels. Also, in the analysis of the mean state, only 11 models were used and provides a reasonably sized ensemble for multi-model analysis and to assess model differences. The text has been revised to "Two models in the PlioMIP2 ensemble, CESM2 and CCSM4_Utrecht, are not examined here due to difficulties in regridding to a common grid."

line 127: authors could be more explicit regarding their grading of a "good climate". Does mean model climate agree well with proxy records or with the overall PlioMIP2 simulation ensemble? Is

location of a model well within multi-model ensemble metrics "good"? Please clarify text accordingly.

We have rephrased this line to show that HadCM3 sits close to the multi-model mean in terms of surface temperature change.

The revised manuscript reads "*It simulates a Late Pliocene warming of 2.9°C, close to the MMM (3.2°C), within the model range (5.2-1.7°C) and is in reasonable agreement with proxy reconstructions (Haywood et al., 2020).*"

line 137-138: key here is the use of the LP land-sea mask in comparison to the originally proposed PI land sea mask, am I right? Please explicitly spell this out for clarity for readers not familiar with differences between enhanced and standard PlioMIP boundary conditions.

We have added the follow text clarifying the difference between the standard and enhanced boundary conditions. *"The enhanced boundary conditions contain the full palaeogeographic reconstruction, whereas the standard boundary conditions retain a modern land-sea mask with the exception of closing the Hudson Bay, Bering Strait and Canadian Archipelago. The enhanced boundary conditions were implemented to be more comparable with the other experiments."*

line 143: For clarity, the sentence could be expanded as follows: "leads to a change in the orography and vegetation since PlioMIP boundary conditions do not discriminate between ice sheet extent and height."

The following text has been added to the manuscript to clarify the challenge in separating the orography and ice sheet boundary conditions: "*One limitation of this set up is that it does not fully separate the impact of ice sheets from orography as changing the ice sheets require changes to the orography and vegetation within the ice sheet regions."*

line 146ff: you speak here of >new< HadCM3 simulations. Since HadCM3 was already part of PlioMIP2 (Haywood et al., 2020), it would be important to explicitly describe in which regard these simulations are comparable to the "classical" HadCM3 PlioMIP2 simulations to which they are compared. Were there any changes in HPC, model version, model parameterizations, or spinup procedure between the older and newer simulations that would be relevant to be highlighted in this section?

We have added the following text into the revised manuscript to clarify how the new simulations relate to the classical simulations: *"The experiments were branched off at model year 3500 from fully spun up from HadCM3 PlioMIP2 experiments and ran for a further 3500 to allow for the climate to reach an equilibrium state (measure by top of atmosphere radiation balance). These experiments were run using the same model parameters as the PlioMIP2 run presented in Hunter et al. (2019). As the experiments were branched off from fully spun up runs, and ran to equilibrium, the results are comparable between the forcing factorisation experiments and the pre-existing simulations presented in Hunter et al. (2019)."*

The PlioMIP2 simulations from Hunter et al. (2019) were performed on the ARC3 (the University of Leeds high performance computing system). The new experiments presented in this manuscript were completed on ARC4, a newer, but similar hardware architecture. The runs on ARC3 and ARC4 have been rigorously benchmarked for multiple variables and for serval climates including the Pliocene. There are no differences between simulations on ARC3 and ARC4 beyond those due to internal variability (i.e. differences between ARC3 and ARC4 do not exceed differences between two distinct 50-year periods on ARC3).

line 168-169: Please already name here the different pressure levels that you will be analyzing, and please also provide a motiviation why you use exactly these in contrast to other levels. In particular the link between, and conclusions drawn from, the analysis of jet stream dynamics in higher and lower pressure levels is in my opinion not always clear from your manuscript.

We the following text has been added at the relevant point in the methods section "As in Li et al. (2015), the zonal wind is examined on three pressure levels (850, 500 and 200 hPa) .".
We now include a figure showing the vertical profile of the wind speed in HadCM3 (and CCSM4-UoT and COSMOS in the supplementary) to allow for comments about changes in the zonal wind speeds as lower pressure levels to be made.

Line 170ff: You refer here to the method by Lie et al. (2015). Since that publication is behind a paywall it would be important to provide as many relevant methodological details as possible also in your manuscript. Outlining selection of pressure levels, and reflecting on the definition of the height to find the maximum of the jet stream, could be particularly relevant towards comprehension of your analysis method. Since you focus your analysis on three discrete pressure levels, can you be sure to fully capture changes in jet stream position and strength that may occur at intermediate pressure levels and that may be likewise relevant towards the understanding of the dynamics that your study aims at?

We have clarified the pressure levels used the Li et at. Following the comment here, we looked at the change of zonal wind speeds at different pressure levels in HadCM3. We find that there is minimal change in the vertical position of the maximum zonal wind speed within HadCM3. Furthermore, not all models provide data on all pressure levels, but 200 hPa is common to all models. Therefore, to allow for as close of a comparison as possible we retain the use of 200 hPa for the analysis in the variability of the jet stream. This analysis has been added to the supplementary material along with a reference in the main text stating that this adds some uncertainty into the analysis.

line 179: "but is not done here" - I suggest to provide a (reference to) the motivation why this is not done here.

This manuscript focuses on changes to the jet stream variability in the Northern Hemisphere, specifically the North Pacific. Adding in more regions could complicate the discussion of the results which is why we choose to focus on one region.

line 180ff: you speak here about variability of the jet stream. Maybe more precisely speak of variability in the jet stream's wind speed maximum? There are also other aspects of variability wrt. the jet stream (position, height, width, etc.)

We have changed the text to "*To measure the variability of the jet stream latitude (referred to as jet stream variability), we calculate the standard deviation of the latitude of maximum wind speed in the month of January at 200 hPa*." to reflect what aspect of the jet we are referring to.

line 191-195: One could reflect on the difference in agreement of models and observations between North Pacific (better) and North Atlantic (worse)

These differences have been noted in the manuscript including a reference to general climate model performance in the North Atlantic region.

line 202: to better identify the models that you are singling out here, change text to "exhibiting a large equatorward shift >4°"? I also suggest to add a comma after "The mean"

No reference to the 4 degree shift will be mentioned but the revised manuscript contains more references to the model differences. The comma has been added.

line 208: please check formulation "in to the", add comma after "layer", and check the formulation "indicating that the zonal wind at each layer are linked" - for the latter, beyond grammar, the implications are not fully clear to me. Do the authors want to suggest here that changes in winds at each layer are linked to each other, so that proxy derived evidence could be employed to verify modelled jet stream changes across the atmosphere column? Please clarify.

Examining the vertical profile of the zonal winds in HadCM3 (Figure R1), the change in the jet stream position is consistent through the depth of the atmosphere with a polewards shift at the lower levels in the atmosphere. This will now be included in the manuscript to clarify this point.

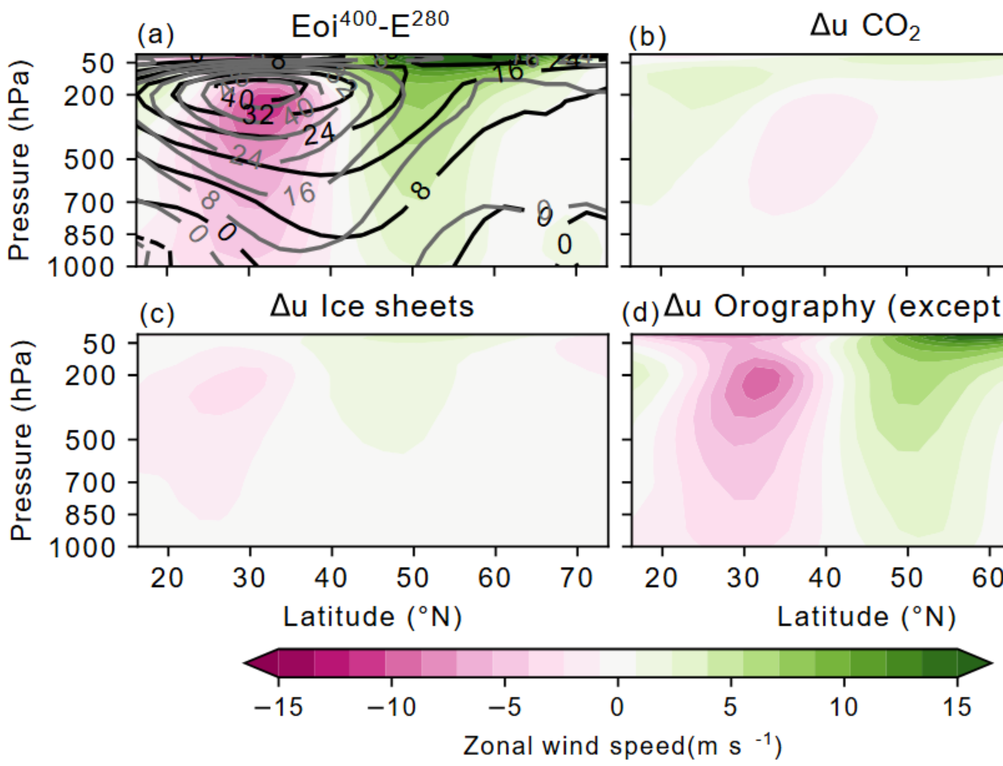

Figure R1 Change in the DJF North Pacific zonally averaged, zonal winds in HadCM3 and the contribution from CO2, ice sheets and orography. The grey contours show the pre-industrial (E280) values and the black contours show the Late Pliocene (Eoi400) values.

line 208ff: This whole section appears to me more a discussion than a result, move it to the respective section? I think it would fit into outlook and further directions since the proxy based information is only touched by the authors, rather than fully explored (in my humble opinion).

As this study contains the results and interpretation (discussion) in the one section this section was included here.

line 219: Please provide reference for the statement that links spatial resolution to jet stream behaviour.

We have added a reference to Athanasiadis et all 2022 (DOI: https://doi.org/10.1175/JCLI-D-21-0515.1)

line 224-225: please be more specific - are they better, could they be better?

This statement was included to reflect the ability of the model to capture Late Pliocene climate by stating that models that more closely represent proxy data, may do a better job so simulating climate in the Late the Pliocene. It is difficult to state if they are better at capturing the behaviour of the jet stream in the Late Pliocene since there is no direct way of representing the jet stream in proxy data. The models listed in this passage do show a weaker and poleward jet stream meaning that it could be possible that models that capture SSTs also do a better job at representing Late Pliocene climate if the Late Pliocene did indeed have a weaker and poleward jet stream (which is what we suggest based on modelling results). This will be clarified in the revised manuscript.

line 245: I think here it is particularly important to replace the term "orography changes" with "gateway changes"; Weiffenbach et al. (2023) do not mention the terms orography or elevation even once, bathymetry is used only once, but then in relation to the description of the HadGEM3 setup rather than to specify details in ocean gateway shape or other bathymetric features

We have rewritten this be more specific. It now reads *"This enhanced AMOC, due to land sea mask changes, is one contributor to the reduced meridional temperature gradient, as more heat is transported northward."*

line 253: extend the sentence to "are inherently linked per boundary condition design in PlioMIP"? If, instead of a combined ice sheet / elevation data set, ice sheet reconstruction was provided via both ice sheet extent and ice sheet thickness, then one could test the impact of albedo vs. elevation effects by means of a "white-blanket" experiment

One could perform an analysis of the impact of the change in elevation and albedo on the jet stream by performing a green mountain and white plain style study (Roberts and Valdes 2017), however that is beyond the scope of this research.

line 263: not sure whether the term "varied" is the best choice here since you talk of factorization with separated variation of boundary conditions; maybe replace "varied" by "diverse"? Furthermore, the sentence reads as if you talked of a variability of a variability - please check and clarify formulation where necessary. I think you may speak here of "The model-to-model discord is larger than the difference caused by the change from PI to LP boundary conditions", but I am not sure.

This section has been rephrased to *"The response to Late Pliocene boundary conditions across models is diverse (Fig 8). The difference between models is larger than the difference caused by a change from pre-industrial to Late Pliocene boundary conditions,"* for clarity

Line 264: consider reformulating the text "no conclusions can be drawn from this data set" (you actually draw conclusions directly thereafter)

We have rephrased this to ' indicating that care should be taken in the interpretation these results as the model error is large' to show that while the model error in the change in variability is large, it doesn't mean the results can't be used, just that care should be taken when viewing the model ensemble and also to consider how each model performs in its representation of the PI/PD jet stream. As in the original manuscript this is followed with a consideration of the models ability to simulate the jet stream in the PI/PD.

line 266: also here orography refers more to land sea mask?

We have kept this as it is in the original manuscript, we hope that clarifying that orography includes land sea mask changes in earlier sections is sufficient.

line 281: orography -> land sea mask?

Here we keep with just the term ‘orography’. This is important in this line as the change in the temperature gradient could also be impacted by a change in vegetation which is included in the orography term.

line 284: fix typo "to c achange"; I do not unerstand the meaning of the formulation "The results of the forcing factorisation hold, independent of the 100 year period chosen to perform this analysis". Please clarify formulation accordingly.

This line was added to indicated that the results of the difference in variability between experiments is not dependent on the 100 years window used to perform the analysis. We have now removed this sentence as it is not necessary to the section

line 289: I do not understand the formulations "split in sea ice" (do you mean difference in sensitivity to sea ice?) and "reduction in Arctic sea surface area" (do you refer here to ice covered Arctic Ocean area?)

Here we are referring to the fact that the experiments with Late Pliocene orography (including land-sea mask) are in a group that all have less sea ice than the experiments with pre-industrial orography (including land sea mask), which leads to two groups of response on the North Pacific jet stream. We then state that the change in sea ice extent could be either due to a stronger AMOC, warming the Arctic and decreasing sea ice cover, or that when using a Late Pliocene land sea mask, some areas of the Arctic ocean are replaced by land, reducing the area of ocean that is available to be covered by sea ice, or a combination of both. If there is less ocean area in the Arctic, then that would reduce the area of sea ice even without a change in the climate. To improve clarity, we have now reformulated this section to *“This grouping in the sea ice could be explained by either a stronger AMOC or a reduction in Arctic sea surface area due to the implementation of LP boundary conditions (or a combination of both)”*.

Line 291-292: Here an explict analysis of polar amplfication vs. jet stream metrics may be useful to support or refute this statement.

The relationship between polar amplification and zonal wind speeds has been examined across the PlioMIP2 ensemble but no strong relationship we found so we do not include it here. It is likely that the changes seen in the zonal winds do not direct correlation with polar amplification as the differences in the model (e.g. resolution) may play a larger factor in determining change to the jet stream than polar amplification.

line 294-295: Indeed, COSMOS is an outlier. Yet, the formulation that you use could suggest that it was the only model for which the jet stream was stronger and less variable in the LP, while in reality there are two more models (so in total 3 of 11 models) showing a similar effect (at smaller amplitude), see Fig. 9. I suggest to clarify the formulation accordingly.

This statement is referring to COSMOS being different from the other two models which completed the forcing factorisation runs (HadCM3 and CCSM4-UoT), the manuscript has been revised to clarify this.

line 304-306: I think here a reference to AABW from the AMOC plots (Fig. 6) my be relevant.

In the revised manuscript, the AMOC plot has been moved to the supplementary as it is only included in the paper to show that the paleogeography is impacting the AMOC. A full analysis about the changes, and drivers in the AMOC is beyond the scope of this study.

line 308: add comma after "set up"? Here I do not fully understand the reference to internal variability of climate models. Do you mean here time-variability (which would be the classical term that comes to my mind when speaking of internal variability), or are you aiming at structural uncertainty as in the PMIP triangle (Haywood et al. 2016), referring to uncertainties in forcings, boundary conditions, their implementation, and model initialization?

Here we are referring to time variability in models, relating to natural unforced variability of the climate system which adds a layer of uncertainty in climate modelling. This is to reflect on the limitation of using 100 years from only one run of each experiment model which may not allow for how the model represents time varying variability or for the period we chose to analyse potentially being out of phase with other models (or a different run of the same model) to be assessed.

line 315: here you explicitly list vegetation as a cause for changes in the meridional temperature gradient; yet, if I understand correctly, its contribution is not explicitly separated from other boundary conditions in your analysis, so its quantitative contribution cannot be related to those by other drivers based on your work alone. Yet, there are previous studies that could be cited. O'ishi and Abe-Ouchi (2013) show that vegetation dynamics reinforce LGM cooling with a particular impact in the northern high latitudes, while O'ishi et al. (2021) and Arima and Yoshimori et al. (2025), and references therein, show a similar reinforcing effect of vegetation changes for warmer climates at orbital time-scales.

In the manuscript we discuss the general impact of vegetation changes on temperature. We have now expanded this section to add the following text *"A change in vegetation can also lead to a large change in temperature, due to different plant types having different impacts on temperature through albedo and evapotranspiration (Bonfils et al., 2012). Studies examining other time periods have reported that changes in vernation in response to climate reinforcing the signal, especially at high latitudes (O'ishi and Abe-Ouchi, 2013; O'ishi et al., 2021). It is likely that the changes seen in HadCM3 are a combination of ocean circulation and vegetation differences between the experiments."*. We do not explicitly look at the impact of vegetation of temperature, rather just argue based on theory that the vegetation change will have some impact, even if we don't have the model runs to explicitly quantify the contribution to warming. To follow the Lunt et al. (2021) forcing factorisation methods, to include vegetation in the forcing decomposition, 16 experiments would be needed which is not possible due to time restrictions with the HPC used.

line 316-317: My understanding is that this rule could be extended - COSMOS has implemented a land sea mask change while nevertheless not showing a weaker jet in LP

While COSMOS (and NorESM1-L) does indeed not display a weaker jet in the Late Pliocene, the reason for this is probably related to the poorer representation of the jet in the models due to a lower spatial resolution. This is not related to the land-sea mask changes which impact the other two models which do not show a slower jet in the Late Pliocene.

line 318: Since you talk about other studies here, please clarify whether with Late Pliocene you refer here to LP as at other locations of the manuscript, or to MPWP, or to the Pliocene more generally.

The revised manuscript has been updated to refer to the 'Pliocene' in the section as the idea of the Pliocene being an analogue from future climate has spanned the use of 'mid-Pliocene', 'mPWP' and Late Pliocene. Therefore, we use Pliocene in the hope that this captures the previous work on the Pliocene as a future climate analogue.

line 329: change "the distribution of key climate variables" to "the key climate patterns"? Regarding the reference to the impact on the model comparison to proxy data, could one or two clear examples be provided for clarity?

We keep the working of 'distribution of key climate variables' as what we mean is how mean state vs extreme values of climate variables change. We have added the line "(for example temperature and precipitation)" which should clarify this.

Tables:

Note, in the revised manuscript, the order of the tables has changed due to changes in the ordering of the text as suggested by reviewers.

Table 1: For clarity one could adapt the last sentence of the caption to: "Models with an asterisk (*) have used for simulation LP a simplified paleogeography with a land-sea mask that is unchanged from modern conditions in the pre-industrial control simulation (E280)." (please check details of implementation in each model once more and adapt formulation accordingly if necessary). Also define abbreviations LP and PlioMIP2 here.

We have adjusted this last sentence as follows "Models with an asterisk (*) use the land sea mask of the pre-industrial control simulation (E280) for the Late Pliocene (Eoi400) experiments."

Table 2: Terminology is a bit unclear. Table 1 lists the various PlioMIP2 models, while you here define PlioMIP3 terminology as the standard. For a study that is largely focusing on PlioMIP2 simulations, and that appears outside the PlioMIP3 context, this may be a bit confusing to readers. For example, when reading the text first I was wondering whether the HadCM3 simulations presented by you already refer to the PlioMIP3 phase. Yet, this does not seem to be the case since you only focus on the overlap of simulations between PlioMIP2 and PlioMIP3. For clarity, I suggest to reformulate the Table header, e.g. as follows: "Boundary conditions from a selection of Pliocene Model Intercomparison (PlioMIP) simulations. Simulations have been prepared in the framework of PlioMIP2 (Haywood et al., 2016, 2020) but are overlapping with the PlioMIP3 simulation ensemble (Haywood et al., 2024). We provide both PlioMIP2 and PlioMIP3 nomenclature for ease of comparison to studies from both frameworks."

We have now expanded the caption to be *"Boundary conditions from a selection of PlioMIP2 runs. These simulations are set up following the PlioMIP2 guidance (Haywood et al., 2016) and overlap with the PlioMIP3 experimental design (Haywood et al., 2024). We include the experiment names for both PlioMIP2 and 3 to allow for comparison between both phases. Here orography also includes vegetation and land-sea mask changes"*.

Figures:

The suggestions for the figures have been implemented into the revised manuscript

Please check font sizes of all text elements in all figures. Some labels and annotations are difficult to read

Figure 1: fix spacing of scalars and units in panel subheadings, define any appreviations PI and LP in the caption

Figure 2: fix spacing of scalars and units in legend; on the right hand side of the x-axis you list Mean, but in the text you speak of Multi-Model-Mean (MMM)

Figure 3: spacing of hPa, definition of time slice abbreviations

Figure 4: spacing of hPa, definition of time slice abbreviations - consider to add "except ice sheets" to the subheader of subpanel d for clarity?

Figure 5: as per my comment to Fig. 4.

Figure 6: Since the AMOC effects are so much dependent on gateway changes, I think towards better understanding the results it would be helpful if the description of each subpanel would be accompanied by listing those settings that differ from modern (gateways, elevation outside ice sheet regions, ice sheets). Furthermore, it seems to me that simulation PI, PI_lp-orog, and LP_pi-ice are favorable to pronounced Antarctic Bottom Water circulation (i.e. the lower limb of the AMOC). These simulations feature modern ice sheets, and, maybe more importantly, modern land sea mask in the Southern Hemisphere, if I am not mistaken. This may be a point to reflect on in your manuscript (and if only as an outlook to the importance of geography in the Southern Hemisphere to AMOC and jet stream dynamics). Potentially reformulate the caption to "Maxima of each experiment annotated." (you use the term "value" twice in one sentence).

In the updated manuscript, this figure will be moved to the supplementary figure as it currently serves to show that palaeogeography is what is causing the difference in the AMOC. As this is not a new finding, we have chosen not to focus on this in this study.

Figure 7: indicate region (North Pacific vs. North Atlantic) shown here in the caption; clarify formulation: "Latitude of maximum zonal wind speed and standard deviation of this latitude (Jet latitude variability)"?

Figure 8: spacing for hPa, definition of time period abbrevations; fix "for in"; clarify choice of January vs. DJF (this potentially in the main text); when trying to understand differences in variability of the jet stream I found this to be difficult based on the metrics provided in the figure. Would it be sensible to also quantify time-variability, as it is present in each panel, as a number, so that different levels of variability can be more easily compared with each other? Potentially you could provide standard deviation over time as an annotation above each subpanel (similarly to done wrt. the AMOC maximum in Fig. 6). Furthermore, but I am not sure due to difficulties with interpretation as outlined above: Is there a link between spatial resolution of a model and jet stream variability evident also from this figure? The models that seem to have the largest variability / most smeared out patterns in Fig. 8 (COSMOS, MRI-CGCM2.3, NorESM-L) also seem to be among the models with lowest spatial resolution (3.75x3.75, 2.8x2.8, 3.75x3.75) - exceptions being IPSLCM5A/2 and MIROC4m that likewise have low spatial resolution but seem to behave a bit different regarding changes in jet stream variability than the other three models.

The mean maximum zonal wind speed and the standard deviation in the latitude of maximum zonal wind speed has now been added to each panel of the figure. A discussion about the variability in jet position and model resolution will be added into the text

Figure 9: Does this figure show results for North Pacific? Clarify caption. Fix typo in "jet stream"; consider adding a comma after "quadrants" and consider deleting the comma after "decreased variability".

Figure 10: spacing in hPa, 100 Januaries (?); also here consider providing a metric of variability that is more easily comparable between simulations (see my comment to Fig. 8). I am asking myself whether it would be sensible to further explore the change in variability due to CO2 in the Pliocene (lowermost two panels). You speak in section 3.3.1 about a small overall contribution of CO2, but it is not clear whether you refer here to a comparison between simulations with modern (uppermost panels) or LP (lowermost panels) geography. I get the impression that there is an impact, but a simple scalar metric like standard deviation (see comments above) could further clarify this.

The figure has been updated to show the mean latitude and the standard deviation in the latitude for each panel. This highlights that there is a small change in the latitude and its variability across experiments with the different CO2 but the same orography and ice sheet configuration however these changes are not large and are smaller than the changes seen between those with differing ice sheets and orography. As these changes are small it cannot be rules out that the internal, unforced variability of the climate model could be contributing to the change seen due to CO2 so no in-depth discussion of this change is included in the manuscript.

Figure 11: capitalize Northern Hemisphere; add information identifying the region and the pressure level for which results are shown here.

Supplementary Material:

The following suggestions have been incorporated into the revised manuscript

Please observe font size in figures (in particular S1, S2, S7, S8, S9, but also other Figures are in some cases at the limit of readability)

Please observe units (no latitude unit provided in S1 and S2 y-axes, no space used in pressure level skalars/units)

Please always make sure that information in figures is self-contained, i.e. that it is clearly explained what is shown and why specific models have been excluded from the analysis / computations:

Fig S1/S2 show three groups of curves without indicating the attribution of each group to a specific quantity. I assume curve groups show results for the three different pressure levels as in the main text. For clarity please indicate them at least in the caption, if not in a legend or as an annotation.

Make clear in captions which season/month is shown, and for clarity use same terminology as in the manuscript (winter->DJF in Fig. S1,2?; Fig. S3,4,6 do not provide any respective information)

Figure S3/4 show >changes< in zonal wind speed, could be indicated for clarity (e.g. "LP-PI zonal wind speed anomaly")

Figure S5 caption is incomplete, I assume it should read: "Change in jet stream latitue vs. ratio of Earth system sensitivity (ESS) and Equilibrium climate sensitivity (ECS). ...". Please also explicitly add a note regarding the reason for excluding HadGEM3 and MRI-CGCM2.3, since the motivation is

not obvious from looking at the figure itself. Further, reference to CCSM4-UoT appears twice in the legend, and I assume this is by mistake.

HadCM3 and MRI-CGCM2.3 were excluded from the regression since they retain a unchanged land sea mask from pre-industrial. This has been noted in the figure caption

Figure S9 caption is incomplete, should it read: "Relationship between January sea ice area and jet stream speed (left), and jet latitude variability (right)". Please explicitly state pressure level(s) for which data is shown here.

Make sure that abbreviations (LP, PI) used in figure elements are defined in the caption for self-consistency (S6, S7, S8). Also define such abbreviations if only used in captions (S1, S2, S3, S4).

Although you provide a lookup table regarding PlioMIP2/3 experiments in the main text, I would avoid mixing terminology (S9 uses old PlioMIP2 naming, while other Figures use PlioMIP3 naming)

When interpreting Fig. S5 in addition to Figs. 7 and 9 of the main text, it would be helpful if the same symbols were used for identical models.

Typos: the blue box groups experiments (S9).

The Following minor comments have all been implemented into the revised manuscript as suggested:

line 54: add the term "meridional" to "temperature gradients"

line 61/62: "linked to" used twice in close proximity

line 99: Do the authors mean to say here: "This will also provide a perspective on the usefulness of the Late Pliocene as a >past< analogue for >future< jet stream variability"?

line 100: LP?

line 143: remove comma after "sheets"

line 152: replace "variable" with "climate variable of interest"?

line 157-158: should reference to Hersbach et al. (2023) be moved to after "ERA5 reanalysis data" for clarity?

line 177: add comma after "December"

line 187: add closing bracket

line 190: weakening of, plus fix brackets

line 198: "being faster" -> "wind speed being higher"

line 242: Please check whether also Zhang et al. (2021) should be cited here

line 226: "To examine possible", delete plural s of jets

line 244: capitalize Northern Hemisphere

line 257: looked at -> investigated?; for clarity, consider adding an "only" after "CCSM4-Utrecht"

line 260: "the the" -> "of the"

line 268: Formulation unclear. Do you mean to say here that "The variability is higher in LP than in PI in most models"?

line 265: when speaking of the change in the jet stream, do you mean here a change in jet stream variability?

line 271: change formulation to "considering jet stream dynamics"?

line 274-275: change formulation to "and higher spatial resolution, indicate that the jet stream is weaker and more variable in the LP"?

line 276: add "(7 of 11)" after "with the majority". This would, in my humble opinion, be a more precise statement that does not conceal that a quarter of the models actually show a different signal.

line 286: I think a reference should be provided

line 290: feedbacks positively -> reinforces?

line 304: than in -> than?

line 310: exhibits -> exhibited?

line 311:: "This" -> "The effect"?

line 312: change to "factors, potentially including model resolution" - if I did not misunderstand that you do not explicitly separate model resolution as a contributor, but you find evidence that suggests that it may play a role.

line 333: add "climate" in front of variability, and maybe provide a few examples (NAO?). This is towards clarity that you speak here of internal modes of variability (ENSO, NAO, AMO, ...) rather than of other aspects of variability