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 will contain a clearer statement on what is being used to define the jet stream. 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 will be 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 will be 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 latitudeand wave breaking will be stated in the manuscript, both in the introduction and in the analysis of the variability. This will reference 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 will 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, and 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 also 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 will now contain a more detailed analysis of the vertical studies of both the zonal winds and temperature. Figure 5 will be replaced with Figure R1 showing the forcing factorization of the vertical temperature profiles. And Figure R2 will be added to show the vertical structure of the zonal winds. Section 3.2 will be 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 will note the similarities between this study and that of Menemenlis et al. (2021) and Oldeman et al (2024). Figure R3 will be included in the revised supplementary material.

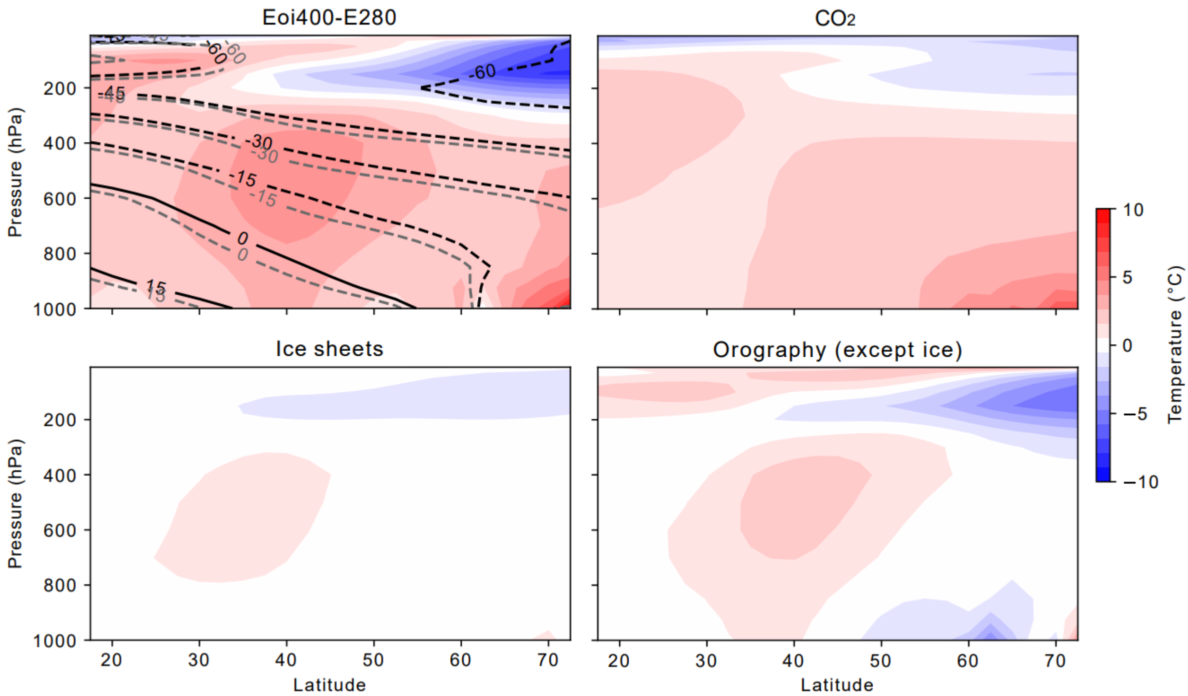

Figure R1: Change in the DJF North Pacific Temperature Profiles in HadCM3 and the contribution from CO2, ice sheets and orography. The grey contours show the pre-industrial (E280) values and the black contours show the Late Pliocene (Eoi400) values.

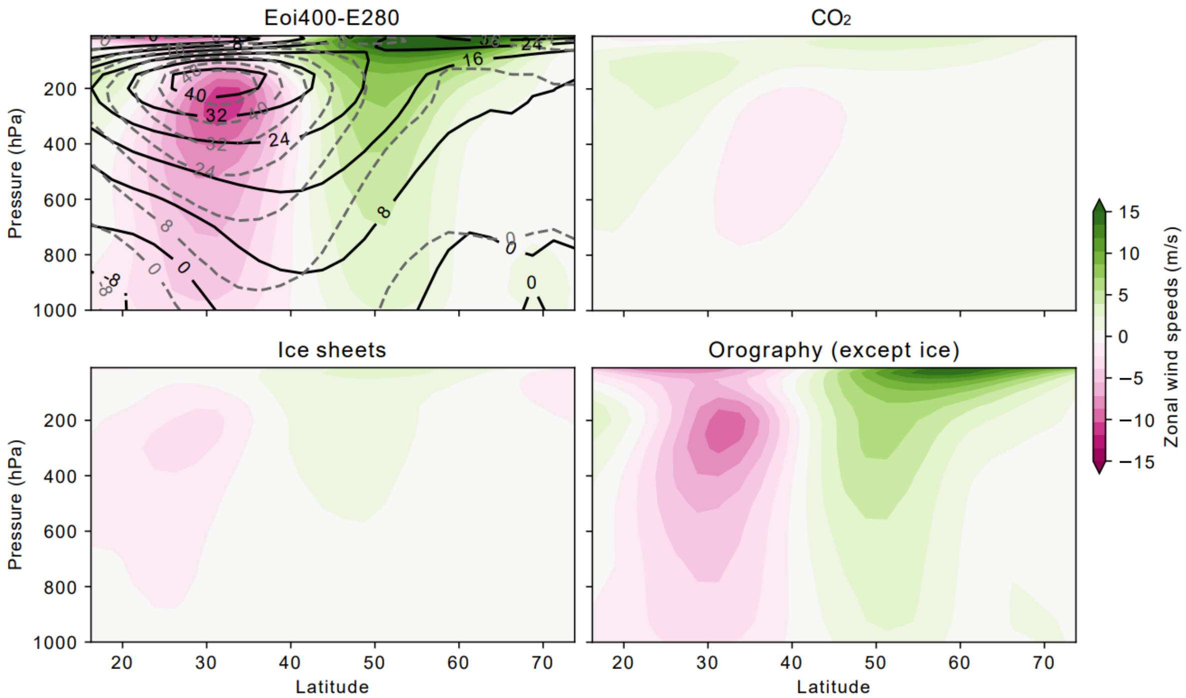

Figure R2 Change in the DJF North Pacific zonally averaged, zonal winds in HadCM3 and the contribution from CO2, ice sheets and orography. The grey contours show the pre-industrial (E280) values and the black contours show the Late Pliocene (Eoi400) values.

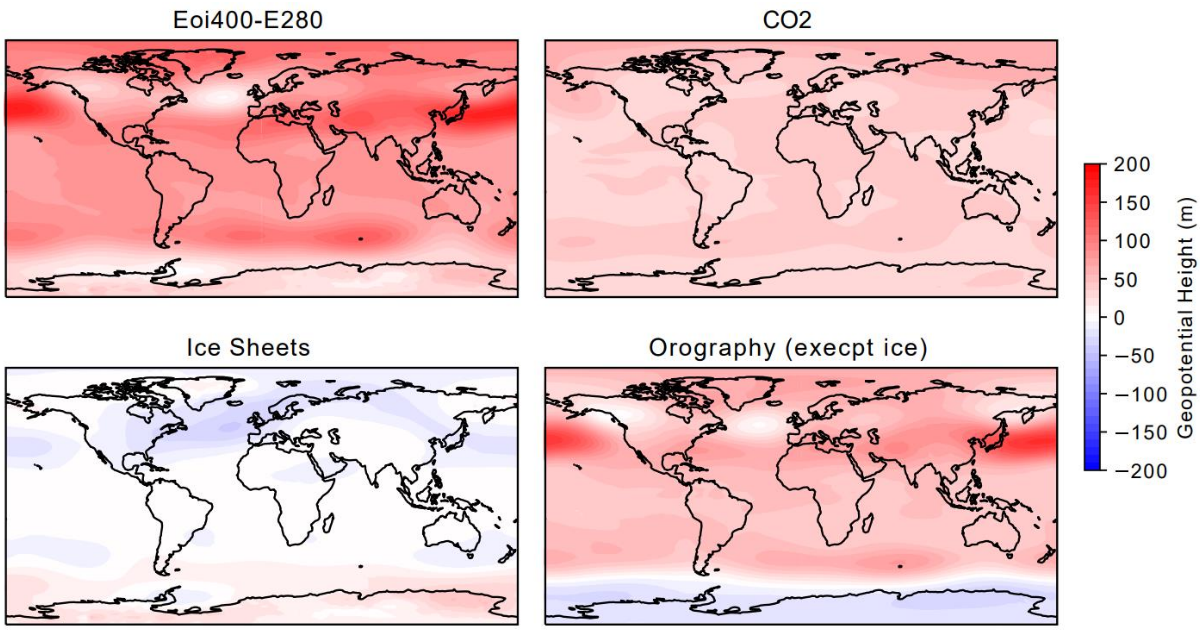

Figure R3 Change in the DJF Geopotential height in HadCM3 and the contribution from CO2, ice sheets and orography.

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 R5) 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. Figures R3 and R4 will be included in the revised supplementary material to show these changes

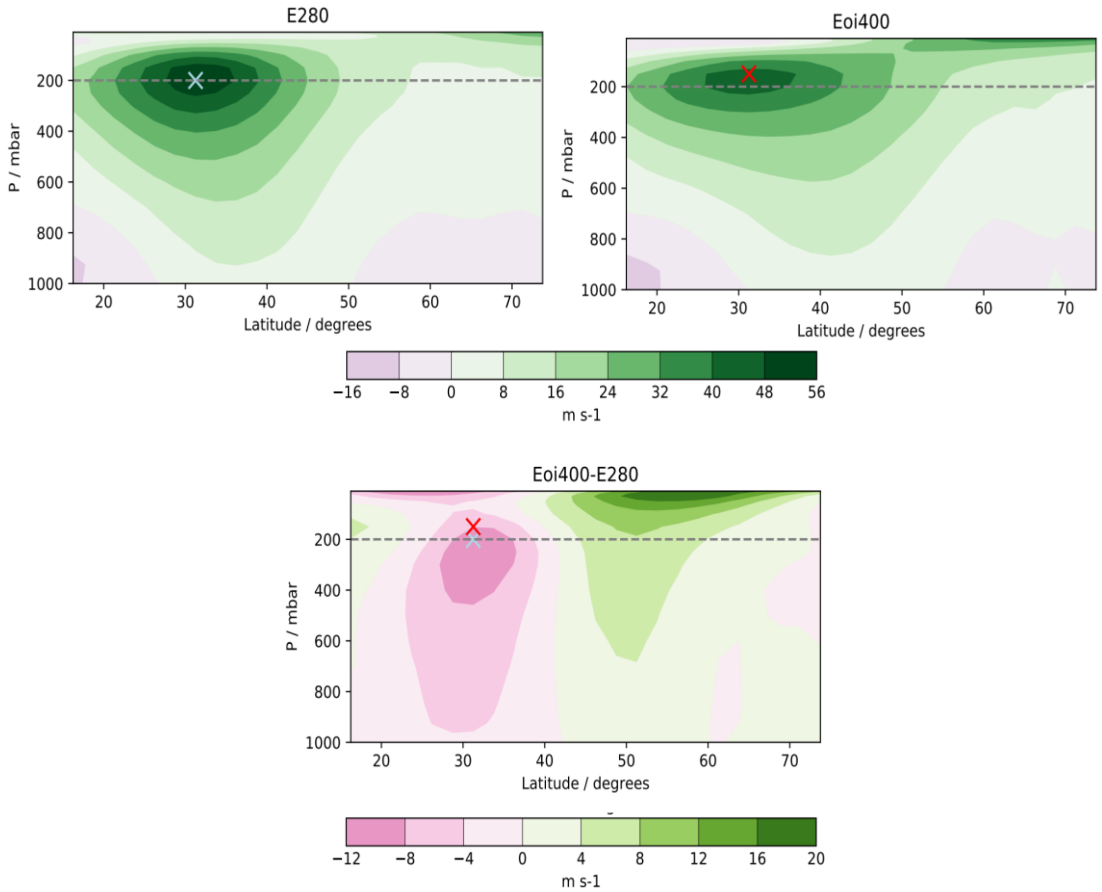

Figure R4: HadCM3 North Pacific zonal wind speeds for Northern Hemisphere winter. The results are presented for the pre-industrial control (E280), the Late Pliocene simulation (Eoi400) and the difference between them. The dashed grey line represents the 200 hPa layer and the blue and red crosses mark the location of the maximum zonal wind speed for E280 and Eoi400 respectively.

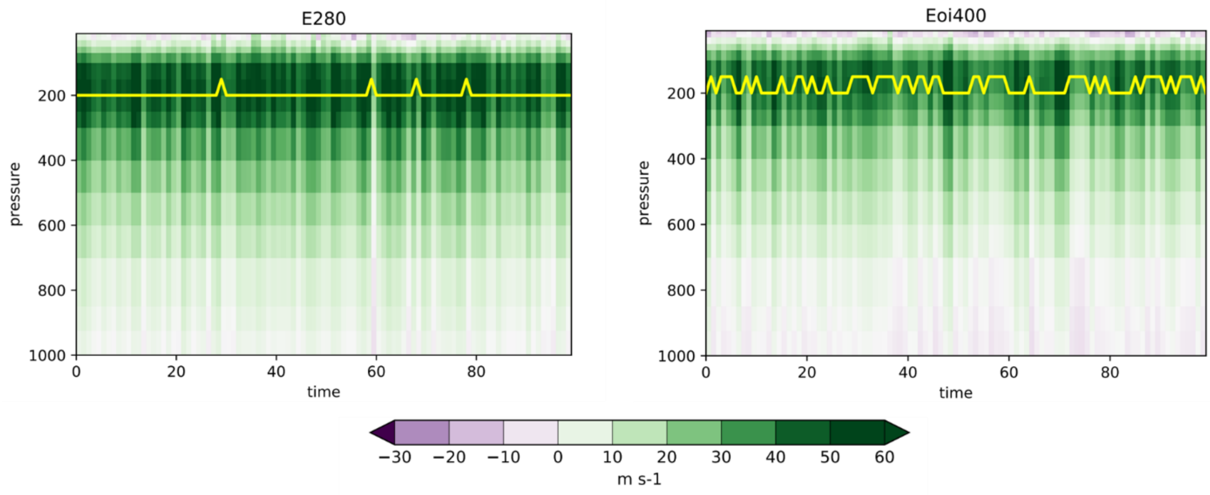

Figure R5: HadCM3 North Pacific zonal wind speeds in the North Pacific (averaged latitudinally between 20N to 40N) for 100 Januarys of the pre-industrial control (E280) and Late Pliocene simulation (Eoi400). The yellow line shows the location of the maximum zonal wind speed.

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.

Subsection 3.3.1 will be restructured to more clearly state the motivation and main findings of the forcing factorisation of the variability and sea ice analysis. 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. In the revised version we will clarify the main message as detailed above.

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 variability could lead to a difference in the distributions of climate variables between the Late Pliocene and Pre-industrial, such as 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 will be 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 will 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 will be updated to avoid the term "good climate". The text will read: "*However, it still produces 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 will include 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 will revise the manuscript 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 will be removed and replaced with a discussion around anticyclonic wave breaking with reference to the Oldeman et al. 2024 paper.

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.

An analysis of the correlation of the zonal wind speeds between pressure levels will be completed and discussed in the revised manuscript.

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 will be revised to contain a discussion about this 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 with 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 will be 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 the geopotential height changes.

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.

The text will be changed to clarity this refers to the temperatures at 200 hPa: "CO_2 causes a cooling in the polar regions, and warming in the tropical regions at 200 hPa which is consistent with observed patterns under modern, CO_2-driven, climate change (Ladstädter et al., 2023)."

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.

Further references to relevant literature will be made to clarify the motivation behind this statement. References will also be made to the similarities of these results with the results presented in Oldeman et al. (2024).

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 9) 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?

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, and the contours have been removed from the figure.

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 assed 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.