Overall, I think this is good paper and should be published in climates of the past with little modification. The topic is appropriate for the journal and quite important in general. I have some suggestions that I think would make the manuscript appeal to a wider audience, which are mainly technical.

AR: We thank Chris Brierley for his helpful comments and positive words. Please find our point-specific responses below.

L35: you refer to a concentration pathway as being an emission scenario. This is confusing, as there is an esmrcp85 variant, but not a esmrcp45.

AR: Yes, we will correct this by removing "emission scenario" from L35.

L97: justify why cosmos is treated differently

AR: COSMOS is treated differently because the curvilinear grid of COSMOS is not suitable for calculating the transports and using the regular 1x1 grid is more straightforward and introduces less error. We will change L97 to: "Ocean freshwater and heat transport have been calculated on native model grids, except for COSMOS where computation is done on a regular interpolated 1x1 degree grid as the native curvilinear grid is unsuitable for this calculation."

L112: What is the magnitude of the errors introduced by using monthly mean temperatures and salinities to compute the potential density and therefore stratification index?

AR: As not all models are able to provide potential density fields, we use the standard TEOS-10 (International Thermodynamic Equation of Seawater – 2010, IOC et al. (2010)) equation of state to calculate the 100-year mean potential density from the 100-year mean fields of temperature and salinity. This method has already been used for other multi-model studies such as Bourgeois et al. (2022) and Muilwijk et al. (2023). Due to averaging effects as well as not all models implementing the same approximation of the equation of state, there may be a small difference with the model-calculated potential density. The magnitude of this difference would vary between models and it is therefore difficult to provide an estimate of its magnitude without having all the model-calculated potential density fields available. We do not believe, however, that the magnitude of error is large enough to significantly impact our results. As the alternative is to not use models that don't have the potential density field available, we consider the possible error to be acceptable.

Table 1: The first 4 models in this table are all variants of CCSM4. You should at least acknowledge this unbalanced ensemble design, and preferably comment on whether this could skew your findings (I suspect not).

AR: There are indeed 5 CESM models included in the PlioMIP2 ensemble. With their respective ECSs of 4.1°C and 5.3°C, CESM1.2 and CESM2 both have a significantly higher ECS than the reported 3.2°C ECS of the CCSM4 models (Haywood et al., 2020), which suggests that the models do respond to forcing in considerably different ways. In addition, the results presented in several published PlioMIP2 studies show that the CESM models all present significantly different results in both the atmosphere and ocean, as can be seen in, for example, Haywood et al. (2020) and Zhang et al. (2021). This is likely related to individual modelling groups using different settings and parameters in their model. For instance, Table 1 in Zhang et al. (2021) lists the vertical mixing parametrization employed by the different modelling groups, which shows that all CCSM4 models use a different vertical mixing parametrization. Therefore, in our opinion, the results would not be skewed by the unbalanced ensemble.

We agree it is important to acknowledge the imbalanced in the ensemble and will add the following sentence to L89: "It is important to note that 5 out of the 15 models in the ensemble are CESM models, of which 3 are CCSM4 models, creating an imbalance in the ensemble. However, due to differences in model versions as well as model settings of individual groups, the response varies widely among the CESM members (e.g. Haywood et al., 2020, Zhang et al., 2021) and we do not expect this imbalance to skew our results."

Fig 1: At no point do you say in this figure caption what temporal averaging is being used. I presume this is annual mean, but showing a winter and/or summer sea ice edge is more conventional.

AR: The temporal averaging is the 100-year annual mean, we will alter the caption to include this information. While a winter and/or summer sea ice edge may be more conventional, we have chosen to only show an annual mean sea ice edge in our analysis because all other variables we present are annual mean variables.

P9: I am very surprised that you are able to discuss AMOC in PlioMIP2 without citing Zhang et al's 2021 paper called "Mid-Pliocene Atlantic Meridional Overturning Circulation simulated in PlioMIP2". This work predates the Weiffenbach et al (2023) paper that is cited on the topic by 2 years! More disturbingly, 13 of the 16 authors of the present manuscript were also authors of the 2021 work. So it seems implausible that you (collectively) are unaware of it. Most importantly, if you (collectively) now have reservations about your (collectively) earlier work then you should be airing them publicly so that the whole community is aware of them.

AR: We did not cite Zhang et al. (2021) due to an unfortunate but not intended mistake. We apologize for any confusion this may have caused. It is indeed important work that should be cited when referring to the stronger AMOC in PlioMIP2. We will refer to Zhang et al. (2021) in addition to Weiffenbach et al. (2023) in L310 and change L180-L181 to: "The MMM AMOC strength increases by 3.1 Sv (+16%) in the mPWP simulations with respect to the PI, which agrees with Zhang et al. (2021) who show a consistently stronger mid-Pliocene AMOC in PlioMIP2."

Fig captions. Please do not start figure captions with uncommon acronyms. When I read any paper, I will first read the abstract and then look at the figures: only afterwards working my way through the main text, if I've found it interesting. I'm sure that I'm not the only person who approaches papers in a similar fashion. I recommend that you try to make your figure captions somewhat standalone. The current reliance on only PlioMIP2 terminology is pervasive. I was also confused by SMM, as even expanding it to "special model mean" does not help explain what the figure is.

AR: Thank you for this suggestion. We will go over all figure captions and make sure to clarify any acronyms in the captions. We think using the PlioMIP2 terminology for the experiments and the SMM in the captions is unavoidable. However, to avoid confusion, we will refer to the Methods section when defining the acronym SMM. We will also add a sentence at the end of each caption to define the experiments (E^{280}, E^{400} and Eoi^{400}).

Fig 10: Consider adding the non-special models (boring models?) to this figure. They would not need to identifiable.

AR: We will add the non-special models to this figure as faded grey circles to provide a more complete figure and explain in the caption that the non-special models are not included in the least-squares fit.

We do want to note that the non-special models are not boring or different from the special models in any climate relevant sense. The only difference from the special models is that they don't have the output data for the E^{400} experiment, which has a mid-Pliocene CO_2 concentration with a pre-industrial orography and other boundary conditions.

L290: Be wary of using "CMIP projections" when I think you mean the scenarioMIP projections. The ice sheet model intercomparison project is part of CMIP.

AR: We will change "CMIP projections" into "ScenarioMIP" projections in L290.

L294: I had forgotten what ~24m relates to, please remind the reader by providing more context.

AR: Thank you for this comment, as it also points out an error in our argument. The ~24 m sea level equivalent decrease in ice volume refers to the sea level increase in the mPWP after reconstructing both the Greenland and Antarctic ice sheet volume. We will correct this error by changing ~24 m into ~21 m, which is the reconstructed sea level equivalent decrease in Antarctic ice volume in the mPWP (Dowsett et al., 2010), and clarify L294 by changing it into:

"It should be taken into account, however, that Antarctic Ice Sheet projections do not show a ~21 m sea level equivalent decrease in ice volume, which is reconstructed for the mPWP Antarctic ice sheet (Dowsett et al., 2010)."

L302: You mention CMIP5 studies here. Has no-one published anything on the topic relating to CMIP6?

AR: Yes, we agree that information on CMIP6 is lacking here. Many CMIP6 Southern Ocean focus on model biases in historical runs, which we discuss in section 2.4.3. We were able to find only one CMIP6 study by Purich et al. (2020) that explicitly discusses future Southern Ocean surface conditions and AABW formation. We will add this reference in the following sentence after L302-304: "CMIP6 models also show future Southern Ocean surface warming and sea ice decrease, as well as mixed layer shoaling that suggests slowdown of AABW formation (Purich et al. 2020)."

L321: You are missing a closing bracket

AR: This will be corrected.

References:

Bourgeois, T., Goris, N., Schwinger, J., & Tjiputra, J. F. (2022). Stratification constrains future heat and carbon uptake in the Southern Ocean between 30°S and 55°S. *Nature Communications*, *13*(1), 340. https://doi.org/10.1038/s41467-022-27979-5

Dowsett, H., Robinson, M., Haywood, A. M., Salzmann, U., Hill, D., Sohl, L. E., Chandler, M., Williams, M., Foley, K., & Stoll, D. K. (2010). The PRISM3D paleoenvironmental reconstruction. *Stratigraphy*, *7*(2–3), 17.

Haywood, A. M., Tindall, J. C., Dowsett, H. J., Dolan, A. M., Foley, K. M., Hunter, S. J., Hill, D. J., Chan, W.-L., Abe-Ouchi, A., Stepanek, C., Lohmann, G., Chandan, D., Peltier, W. R., Tan, N., Contoux, C., Ramstein, G., Li, X., Zhang, Z., Guo, C., … Lunt, D. J. (2020). The Pliocene Model Intercomparison Project Phase 2: Large-scale climate features and climate sensitivity. *Climate of the Past*, *16*(6), 2095–2123. https://doi.org/10.5194/cp-16-2095-2020

Muilwijk, M., Nummelin, A., Heuzé, C., Polyakov, I. V., Zanowski, H., & Smedsrud, L. H. (2023). Divergence in Climate Model Projections of Future Arctic Atlantification. *Journal of Climate*, *36*(6), 1727–1748. https://doi.org/10.1175/JCLI-D-22-0349.1

Purich, A., & England, M. H. (2021). Historical and Future Projected Warming of Antarctic Shelf Bottom Water in CMIP6 Models. *Geophysical Research Letters*, *48*(10), e2021GL092752. https://doi.org/10.1029/2021GL092752

Zhang, Z., Li, X., Guo, C., Otterå, O. H., Nisancioglu, K. H., Tan, N., Contoux, C., Ramstein, G., Feng, R., Otto-Bliesner, B. L., Brady, E., Chandan, D., Peltier, W. R., Baatsen, M. L. J., von der Heydt, A. S., Weiffenbach, J. E., Stepanek, C., Lohmann, G., Zhang, Q., … Abe-Ouchi, A. (2021). Mid-Pliocene Atlantic Meridional Overturning Circulation simulated in PlioMIP2. *Climate of the Past*, *17*(1), 529–543. https://doi.org/10.5194/cp-17-529-2021