Response to Anonymous Reviewer 2- Reviewer comments in black, author comments in blue

Holmes et al. compare observed Antarctic sea ice area trend with the trends in CMIP6 simulations. The authors find that during 1979-2018, the models are not consistent with the observations, as has been widely noted previously. This is because the Antarctic sea ice area tends to steadily decline in response to anthropogenic forcing in GCMs, whereas the observed Antarctic sea ice area increased during the first 35 years or so of the satellite record (roughly 1979-2015). However, strikingly low Antarctic sea ice areas were observed in 2022 and 2023. Due to this, the authors find that when the time period over which the trends are computed is extended to 1979-2023, the models and observations are fairly consistent. The authors conclude that this gives a greater level of confidence in the CMIP6 Antarctic sea ice simulations than previously though.

The manuscript is clearly and concisely written, and the presentation is polished.

We thank the reviewer for their comments and we are glad they found the manuscript concise and readable.

However, I do not find the conclusions to be compelling. The issue is that although we often use linear trends to compare models with observations, the observed Antarctic sea ice area evolution looks strikingly unlike a linear trend plus any straightforward type of noise. This can be readily seen by looking at any time series of observed Antarctic sea ice area or extent (e.g., https://zacklabe.files.wordpress.com/2023/12/nsidc_sie_timeseries_ant_anomalies-5.png). I do not see any meaningful similarity between the steady long-term decline of Antarctic sea ice area in typical CMIP6 simulations (e.g., Historical and then SSP5-8.5 simulations plotted in Fig 4c,d of Roach et al. 2020) and the observed gradual expansion followed by a short-lived abrupt loss during the past two years, even if both time series have similar OLS linear trends.

In my opinion, it is misleading to use the similarity between these linear trends to conclude that "we should now have some level of greater confidence" in the model simulations and that therefore "projections of substantial future Antarctic sea ice loss may be more reliable than previously thought" (quotes from the manuscript). I am not sure what meaningful information can be gleaned from noting this similarity in linear trends between the observations and GCMs during 1979-2023.

To this end, the authors do not include any plots of actual time series, only reporting the OLS linear trends, so a reader unfamiliar with the observations and simulations may be substantially misled by this manuscript. (I don't mean to imply that the authors are being intentionally misleading in any way.)

We agree with the reviewer that "the Antarctic sea ice area evolution looks strikingly unlike a linear trend plus any straightforward type of noise" and they make a valid point that we did not present any of the modelled timeseries in the paper. However, on examining the model data we disagree that the typical modelled behaviour is 'steady decline'. The steady decline that is most obvious in the Roach et al (2020) figure cited is the multi-model mean, which averages out the internal variability and so reflects the gradual forced anthropogenic trend. If we examine all the ensemble members individually (Figure R5), a linear decline is not in general the most prominent feature of variability. Instead the individual realisations have a wide variety of different variability, including some with limited trend followed by rapid decline (e.g. CNRM_CM6_4, HadGEM3_GC31_LL_1), as seen in observations.

Figure R5: 1979-2023 annual mean sea ice area in observations (Sea Ice Index v3, top left) and in all CMIP6 model ensemble members considered in the analysis, sorted by their linear trend over this period. Linear trends are shown and indicated in red (significant at $p<0.05$) or grey (insignificant). Each panel includes annotation showing the simulation's 1979-2023 climatology and trend.

This wide range of behaviours further justifies the linear trend metric; without a hypothesis of what shape we expect sea ice evolution to take, the linear metric is the best 'first test' of the data. We use it openly as a simple 'top level' metric, which is often used, and our over-arching purpose is to demonstrate that model-observation consistency 'cannot be ruled out on trends alone'. Our argument is that IF we only consider linear trends of annual-mean ice area THEN observations recently came into line with models. We think that with appropriate discussion this is a valid and noteworthy conclusion.

INTENDED ADJUSTMENTS:

- Add a note to the discussion describing that an OLS linear trend is a limited parametric assessment of a timeseries, and the observed time series does not particularly resemble a linear trend with noise imposed. However, while the multimodel mean resembles smooth decline (e.g. Roach et al 2020), individual members display complex behaviours which more clearly show the interplay of trends and variability. Therefore, while the linear metric gives a first indicator of model performance, this further highlights the need to probe higher-order or non-parametric characteristics of modelled sea ice variability to investigate the models' fidelity.
- Add an Appendix 'Timeseries' which will contain Figure R5.

Furthermore, the authors mention that the observations are consistent with "the 'two-timescale' response to stratospheric ozone forcing whereby increasing westerlies cause a sea ice increase on 'short' timescales and decline on 'long' timescales (Ferreira et al., 2015; Kostov et al., 2017)." It should be emphasized that a later study that included some of the same authors (Seviour et al. 2019, doi:10.1175/JCLI-D-19-0109.1) concluded that this ozone forcing mechanism is unlikely to be a primary driver of the mismatch between the Southern Ocean surface cooling (and Antarctic sea ice expansion) in observations and the Southern Ocean surface warming (and Antarctic sea ice retreat) simulated by GCMs.

We agree that we need to edit the text here. We have re-visited the Seviour et al. (2019) manuscript, which indeed shows that (based on relationships between model climatology and the cooling and warming phases of the SST response), the two-timescale response to ozone depletion is unlikely to be a primary driver of the model-observations mismatch. We do not think this precludes our mentioning this as a possible hypothesis, but it merits a more nuanced discussion.

INTENDED ADJUSTMENTS: We will add a note to the manuscript to confirm that this is a possible hypothesis and referencing the Seviour et al. (2019) paper, and clarify 'models under-estimating the 'two timescale' response… " to 'models under-estimating the timescale or magnitude of the cooling phase of the two timescale response'.