Brief Communication: ~~Antarctic sea ice loss brings observed trends into agreement with climate models~~New perspectives on the skill of modelled sea ice trends in light of recent Antarctic sea ice loss

Caroline R. Holmes[1], Thomas J. Bracegirdle[1], Paul R. Holland[1], Julienne Stroeve[2,3,4], Jeremy Wilkinson[1]

[1]British Antarctic Survey, Cambridge, CB3 0ET, UK
[2]National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA
[3]Centre for Earth Observation Science, University of Manitoba, Winnipeg, Manitoba, Canada
[4]Earth Sciences, University College London, London, UK

Correspondence to: Caroline R. Holmes (calmes@bas.ac.uk)

Abstract. Most climate models do not reproduce the 1979-2014 increase in Antarctic sea ice cover. This was a contributing factor in successive Intergovernmental Panel on Climate Change (IPCC) reports allocating low confidence to model projections of sea ice over the 21st century. We show that recent rapid declines bring observed sea ice area trends back into line with the models, and~~but~~ confirm that discrepancies exist for earlier periods. ~~This implies that projections of substantial future Antarctic sea ice loss may be more reliable than previously thought, with wide-ranging implications for the evolution of the Southern Hemisphere climate~~ This demonstrates that models exhibit different skill on different timescales. We discuss possible interpretations of this~~the~~ ~~changing~~ linear trend assessment given the abrupt nature of recent changes, and the implications for future research.~~judgement,specifically whetherwe can now have .~~

1 Introduction

The early years of the twenty-first century revealed a puzzling conundrum in Antarctic sea ice (Turner and Comiso, 2017; National Academies of Sciences and Medicine, 2017). Observations of Antarctic sea ice extent (SIE) showed a small increase during the satellite era (which began in late 1978), with annual mean values reaching a maximum in 2014, but most climate models simulated SIE declines over the same period. Various studies examined possible reasons for this discrepancy (Turner and Comiso, 2017). Specifically, the community discussed whether it could be explained by internal variability masking the anthropogenic forced signal in observations (Gagné et al., 2015; Rosenblum and Eisenman, 2017; Roach et al., 2020) and the extent to which it revealed model deficiencies in sea ice processes (Fox-Kemper, 2021). Some studies found that the observed pan-Antarctic trends lay within the distribution of modelled trends (Polvani and Smith, 2013; Zunz et al., 2013) and that only regional trends could robustly be deemed inaccurate in the models (Hobbs et al., 2015). However, these studies considered trends to 2005 only, and over this 27-year period the role of internal variability is larger than it is for longer periods with more recent end dates. Others suggested that trends in sea ice, particularly SIE, may not be a robust metric of model performance, particularly when the observational time series is too short to separate internal variability from anthropogenic forcing (Notz, 2014). Even so, the poorly understood discrepancy between models and observations has been a contributing factor in a widespread lack of confidence in projections of 21st century Antarctic sea ice decline, and consequently in many aspects of projected climate change around Antarctica~~,~~ which are underpinned by projections of substantial sea ice decline (Bracegirdle et al., 2015; Bracegirdle et al., 2018).

Recently, Antarctic sea ice has exhibited a starkly different pattern of behaviour. Following the pre-2015 era of slightly increasing ice extent, rapid ice loss beginning in early 2015 culminated in a dramatic drop in spring 2016-17 (Turner et al., 2017). This led to several years of record low SIE, which has been framed as a ‘new sea ice state’ (Purich and Doddridge, 2023; Hobbs et al, 2024). This situation shows no sign of abating, with further declines since 2021 leading to monthly-mean SIE records being broken in eight months of 2023 (Fetterer, 2017; Siegert et al., 2023). The initial decline showed strong linkages to patterns of intrinsic atmospheric variability (Turner et al., 2017; Schlosser et al., 2018; Zhang et al., 2022) which have high internal variability on short (sub-annual) timescales. However, growing evidence of the contribution of warming in the subsurface ocean (Zhang et al., 2022; Purich and Doddridge, 2023), and the magnitude and spatial homogeneity of the sea ice reductions since 2016/17, point to more sustained declines.

We are therefore interested in the fundamental question as to whether this new data showing rapid decline should change our judgement of the models’ skill. ~~In light of this sea ice loss,~~ To do so, in the context of previous assessments and based on the approximate linearity of the modelled time series, we assess linear trends. Specifically, we re~~-~~consider whether the distribution of linear trends simulated by the current generation of climate models, from the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016) dataset, allows for a trend of the observed magnitude and thus whether observed trends are consistent with the multi-model ensemble. Key previous studies have considered trends to 2005 (Hobbs et al., 2015; Polvani and Smith, 2013; Zunz et al., 2013) or 2013 (Rosenblum and Eisenman, 2017) based on CMIP5 models and to 2018 based on CMIP6 (Roach et al., 2020). We might expect the situation to have changed, for two reasons. First, being able to assess trends in longer timeseries (due to the longer observational record) potentially reduces the impact of short-term internal variability on trend calculations (Notz, 2014). Second, and more specifically, these data now include the recent years of observed rapid decline of sea ice, decreasing long-term trends. Therefore we perform an analysis of all trends with end dates between 2005 and 2023, to place our results in the context of previous studies and show how the results change over time due to these two factors, while using a consistent set of CMIP6 model data (such that the changes are not attributable to changes in model components or resolution). Our discussion of these results focusses on the changing assessment of skill depending on the timescale considered, the implications for our confidence in the models, and the interpretation of linear trend assessments considering the abrupt nature of recent changes.

2. Data and Methods

2.1 Sea Ice Metric

Sea ice cover is calculated as either sea ice extent, SIE (the total area of all gridboxes where sea ice concentration SIC exceeds a 15% threshold), or sea ice area, SIA (the sum of gridbox areas multiplied by gridbox SIC). SIA has larger observational uncertainties, as it is more sensitive to differences in SIC. However, SIE is a non-linear measure and so can give misleading results when comparing models and observations or when calculating trends (Notz, 2014). Therefore, in contrast to some previous assessments, but following community precedent (Roach et al., 2020), we assess SIA. SIA and SIE have similar trends (Fig. A1).

2.2 Model Data

We use data from 39 CMIP6 models, from multiple modelling centres. Across the ensemble, there are multiple different model components and resolutions of each component. Monthly SIA is obtained from the University of Hamburg (UHH) CMIP6 Sea Ice Area directory (https://www.cen.uni-hamburg.de/en/icdc/data/cryosphere/cmip6-sea-ice-area.html, accessed 2023-08-17) and aggregated into weighted annual means. This is supplemented by SIA for the two NorESM models, which are not available in the UHH dataset due to a bug in an earlier version of NorESM released SIA data. We merge historical simulations ending in December 2014 with the ssp585 forcing scenario run for 2015 to 2023. ssp585 indicates a global average radiative forcing of 8.5 W m^{-2} by 2100 (O'Neill et al., 2016). This is a high-emissions forcing scenario; however, emissions scenarios have little bearing on results for the time period considered here. The resulting historical-ssp585 merger constitutes 188 ensemble members from 39 models (Table A1), each contributing between 1 and 57 members of an initial condition ensemble.

By using a large number of ensemble members of the historical multi-model ensemble, we sample internal variability under historical anthropogenic forcing. However, since only four models contain more than six members, we use a maximum of six members from each model to avoid weighting the results too heavily towards models with large ensembles. Thus the final ensemble analyzed has 98 members (Fig. B1) from 39 models (Table B1). The sensitivity of our results to this treatment of model ensembles and to the emission scenario is discussed in Appendix C.

Since many models have drifts in their pre-industrial runs, we calculated linear trends over the full pre-industrial period available (in the range 150 to 500 years across the 32 models with data available in the UHH dataset; Table B1), henceforward referred to as ‘drift’. In all cases, drifts are an order of magnitude smaller than the trends for years 1979-2023, and there is no significant inter-model relationship between the drift in a model’s pre-industrial simulation and the ensemble mean of linear trends in that model (p=0.48). This implies drifts are negligible in the context of historical trends, consistent with results for CMIP5 (Gupta et al., 2013), and so they are not considered further.

2.3 Observational Data

For an estimate of observed sea ice cover, NSIDC Sea Ice Index v3.0 SIA (Fetterer, 2017) is used, available from January 1979-September 2023. We investigate the role of observational uncertainty by also using other observational estimates for 1979-2019 from the UHH SIA dataset (Dörr, 2021). Data for missing months (December 1987-January 1988 for the Sea Ice Index v3.0) are infilled by interpolating between the same month in the previous and following year (Rosenblum and Eisenman, 2017).

2.4 Trend evaluation methodology

Our evaluation methodology is an extension of that previously used for CMIP5 (Rosenblum and Eisenman, 2017). Linear trends are calculated for all periods of at least 35 years overlapping with the satellite record (January 1979-September 2023) using the OLS method of the Python package statsmodels.api. For comparison with the earlier studies mentioned in the Introduction, we additionally calculate trends for periods 1979–y2 where y2 is between 2005 and 2012. We calculate the mean and standard deviation of the trends from the model ensemble and use these to fit a Gaussian distribution, with cumulative distribution function F(X), to the distribution of modelled these trends. To estimate the probability that a trend at least as large as observed would occur in the climate model population, we calculate the p-value for a one-tailed test as 1-F(x), where x is the observed trend. The extent to which a linear trend is an appropriate metric for evaluating SIA, given the evidence for a recent regime change, is considered in the Discussion below.

3 Results

3.1 Trend evaluation

The recent decade of data has reduced the significant positive trend (Parkinson, 2019) in observed annual-mean and monthly SIA, which peaked in the period ending 2015, to near-zero (Fig. 1a)-c), red lines; Fig. C1a, A1). For some months and in the annual mean, the trend since 1979 is now weakly negative, and trends are statistically insignificant in all months (Fig. A1). Meanwhile, adding the extra years of data hardly changed the multi-model mean trend at all (Fig. 1a)-c), blue lines). The mean trend remains strongly negative, although a few simulations have weakly positive trends. The simulated trends are less influenced by internal climate variability as more years are added, and therefore the standard deviation of the modelled trends for a fixed start year of 1979 decreases over time (Fig. C1c).

In light of these findings, we test the null hypothesis that observed sea ice trends are consistent with trends simulated across the CMIP6 multi-model ensemble, and consider how additional years of data affect the outcome of this test. We consider trends calculated with both a fixed start date (1979) and fixed duration (35 years) to aid our interpretation. Until 2010 inclusive, the probability of a CMIP6 model trend matching or exceeding the observed trend exceeds 0.05, so we would ~~accept~~ not reject the null hypothesis that modelled and observed trends are consistent (as concluded in Zunz et al., 2013; Hobbs et al., 2015;

Polvani and Smith, 2013)). However, in the period 2005 through 2015, the multi-model mean trend and observed trend diverge while the modelled trend distribution narrows (Fig. C1), reducing the likelihood that the observed trend falls within the modelled distribution. As a result, between 2011 and 2018, the probability of a CMIP6 model trend matching or exceeding the observed trend is very low ($p<0.05$; Fig. 1d), so the null hypothesis is rejected and the model trends may be deemed inconsistent with observations. This test provides a clear result; the short time period of under forty years should allow for a generous range of modelled trends due to internal variability, but this range still fails to accommodate the observations.

From 2015, the probability of CMIP6 trends matching or exceeding the observed trend starts to increase, as the ice loss brings observations into line with the models (Fig. 1d). However, if trends are calculated with a fixed 1979 start date, progressively lengthening the trend under consideration decreases the modelled trend standard deviation while hardly affecting the model mean trend (Fig. C1). This makes it less likely that the observed trends will fall within the distribution of modelled trends.

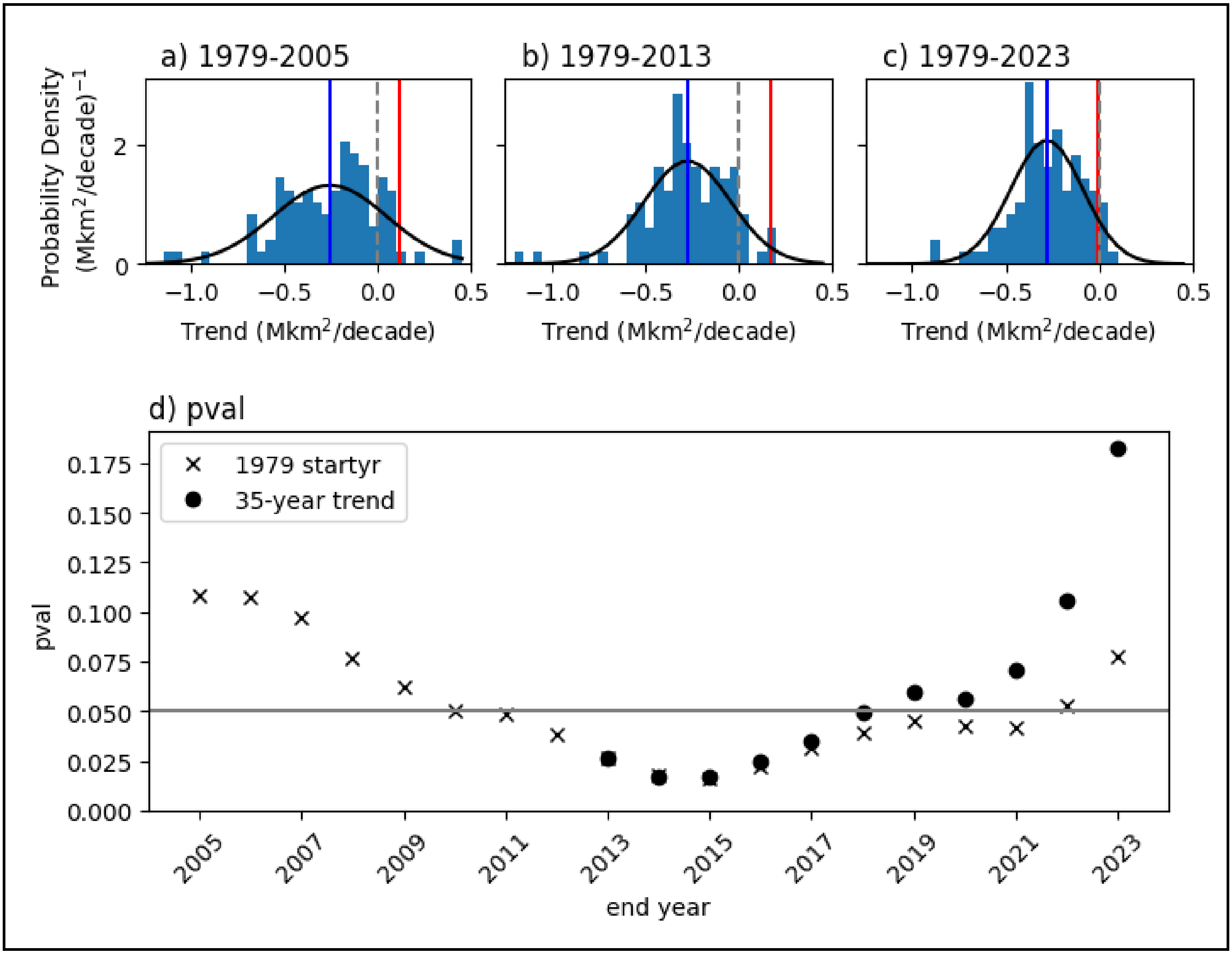

Figure 1 (a-c) Linear trends in annual mean SIA in satellite observations (red) and CMIP6 models (blue histogram) and Gaussian fit to CMIP6 distribution (black) for the periods (a) 1979-2005, (b) 1979-2013 and (c) 1979-2023. The dashed vertical line indicates zero trend and the blue line indicates the multi-model mean. (d) the probability of observing a trend at least as large as observed (a one-tailed test) under the null hypothesis that observations are taken from the same population as the CMIP6 multi-model ensemble, for varying end dates and either a fixed start date of 1979 as in panels a) and b) (crosses) or fixed trend length of 35 years (dots).

Only in 2022 does the recent rapid decline in observations counteract this effect and finally bring observed trends into line with the models (null hypothesis not rejected at p=0.05; Fig. 1d). In contrast, for ‘fixed duration’ trends, the standard deviation of modelled trends remains large, while the observed trend more rapidly declines and becomes negative due to the neglect of early low SIA years (Gagné et al., 2015; Schroeter et al., 2023) in addition to the inclusion of the recent low SIA years. Therefore, the null hypothesis is no longer rejected at p=0.05 as early as 2019 under this measure.

3.2 Relationship of trends with mean state

It is known that, seasonally and especially in summer, there is a relationship between sea ice area climatology and future trends, which is to be expected as, for example, very low sea ice constrains trends (Holmes et al, 2022). Therefore, we investigated the relevance of this for our trend assessment. The relationship between both summer and annual mean climatology and the annual mean trends is highly statistically significant, but has a very weak slope (Fig C2a,b). Since there are two models (from the MIROC family) which are clear outliers, having far too little sea ice in the annual mean (Fig C2b; Shu et al, 2020), we test the sensitivity to removing these models. This does not change our conclusion that trends are consistent for an end date of 2023 (Fig. C2c). Therefore, while there is some evidence that the models with trends closest to observations tend to be biased low (Fig. C2a,b), this does not appear to dominate our conclusion that observed and modelled trends are now consistent.

4 Discussion and Conclusions

~~**we have, the more**~~ Our results show that, if we consider linear trends in models and observations, then we find that the level of agreement varies over time. ~~that the consistency between observed and modelled trends changes over time.~~ Firstly, for early end dates (prior to 2011) there is no evidence of inconsistency between observed and modelled trends, as noted by earlier studies (Hobbs et al., 2015; Zunz et al., 2013; Polvani and Smith, 2013). Secondly, there is a mismatch between observed and modelled trends for the period up to around 2018, as discussed in the Introduction.~~, s~~ This suggest~~sing~~ that modelled anthropogenic trends are too strong relative to modelled variability during that period. Finally, our study shows the novel result that the persistent low Antarctic SIA of 2022 and 2023 brings observed trends back into line with the ensemble of modelled trends. Moreover, trends on the shorter 35-year timescale also fall within the model ensemble for the five most recent 35-year periods (Fig. 1d).~~, permitting the interpretation that modelled forced trends and variability are realistic on 45-year timescales (the full length of the modern satellite record). This is an important conclusion, since these longer timescales are of greatest~~

~~relevance to centennial projections of climate change, and to the attribution of anthropogenically forced change. Moreover, even trends on the shorter 35-year timescale fall within the model ensemble for the five most recent 35-year periods (Fig. 1d).~~

~~Focussing on trends with a fixed start date, the changing assessment of skill with increasing years of data could be explained in several ways.~~We approach our interpretation of the changing assessment of skill as follows: c~~C~~onceptually, for any time period there is a distribution of model trends and also a distribution of possible real trends that could have occurred (depending upon the evolution of internal climate variability). The observed trend is a single realisation of the distribution of possible real trends. The observed trends with end dates between 2011 and 2021 were outside the model trend distribution. Now, the latest observed trends fall within with the distribution of modelled trends, as do observed trends for periods ending before 2011. In other words, the observed trends over the middle period lay in the region where the modelled and real trend distributions did not overlap, and observed trends in the earlier and most recent periods lie in the region where they do overlap.

The non-overlapping region could arise from a difference in the spread of the ~~The~~ modelled and real trend distributions (due to ~~will differ in their spread if the models have~~ inaccurate modelled variability) or~~, and~~ in their mean (due to a too-strong~~if the models have an inaccurate~~ modelled anthropogenic forced trend). Therefore, inaccurate variability, particularly on multidecadal timescales, could explain the changing assessment of skill. Indeed, modelled variability exceeds observed variability and varies greatly between models (Zunz et al., 2013, Roach et al., 2020, Diamond et al., 2024), with some models containing large centennial variability (Zhang et al., 2019). Alternatively, it could be that the modelled anthropogenic trends are too strong (Schneider and Deser, 2018), or emerge too early. For example, this is consistent with the hypothesis that models under-estimate the timescale or magnitude of the cooling phase of the 'two-timescale' response to stratospheric ozone forcing, whereby increasing westerlies cause a cooling (sea ice increase) on 'short' timescales and warming (decline) on 'long' timescales (Ferreira et al., 2015; Kostov et al., 2017). However, other evidence from models suggests this mechanism is unlikely to be a primary driver of the model-observation mismatch (Seviour et al., 2019).

We can then consider what our results imply for our question as posed in the Introduction, namely whether recent rapid declines observed in satellite data change our judgement of model skill, and ultimately our confidence in the models. This paragraph considers the answer to this question based on the linear trend assessment, and the following paragraphs take the broader view of how a linear trend assessment should be interpreted in the light of the possible step change nature of recent decline. Our results permit~~ting~~ the interpretation that modelled forced trends and variability are realistic on 45-year timescales (the full length of the modern satellite record). ~~This is an important conclusion, since these longer timescales are of greatest relevance to centennial projections of climate change, and to the attribution of anthropogenically forced change.~~However, the existing discrepancy on shorter time scales points to fundamental issues remaining. If this discrepancy is, as discussed above, linked to multidecadal variability or to ozone forcing, then one interpretation may be that we can have some level of greater confidence in projections of substantial centennial decline (Roach et al., 2020, Holmes et al, 2022) under strong forcing, since model performance on longer (45-year) timescales is of greatest relevance to centennial projections of climate change. However, our confidence would remain low under weak forcings or in the near term, where multidecadal variability and ozone forcing retain relative importance. If, however, the discrepancy is because the forced greenhouse gas response is too strong, models will produce too-strong ice loss even on centennial timescales. Confidence in which of these interpretations is most appropriate will require both more years of data and further analysis. Further, processes lacking from models, such as increasing freshwater input from accelerating ice sheet melt (Swart et al., 2023), may provide further complications in the relative evolution of modelled and observed sea ice over the 21st century.

This study uses linear trend analysis as a metric for evaluation. ~~The importance of our results is in showing that we can no longer rule out climate model simulations of Antarctic sea ice based on linear trends alone.~~Linear trends are a limited parametric assessment and the observed time series when the years 2017-2023 are included arguably looks strikingly nonlinear in time (Fig B1). Indeed, the recent abrupt change has been interpreted by some as a regime shift (Purich and Doddridge, 2023; Hobbs et al, 2024), which points to limitations of applying a linear trend evaluation. Nevertheless, an update to the linear trend evaluation has significant value. Firstly, the use of linear trends in many previous assessments (as cited in the introduction) merits a careful examination of whether the conclusions of those studies still hold. Secondly, many models have approximately linear evolution in time (Fig B1), which justifies a comparison of linear trends, although the time evolution of SIA in many models also exhibits nonlinear features so that the apparent observed nonlinearity itself is not a reason to conclude a discrepancy between models and observations. Thirdly, a regime shift is not the only interpretation of observations, and multidecadal variability superimposed on a forced linear trend (e.g. Zhang et al, 2019) could cause abrupt change as seen since 2016. This interpretation is consistent with evidence of steady sea ice decline in the 20th century before the satellite era (Fogt et al., 2022), and with early satellite data which suggest that the ice area was more variable in the 1960s (Meier et al., 2013; Gallaher et al., 2013) and dropped rapidly immediately before the onset of continuous coverage in 1979 (Cavalieri et al., 2003). In this case, evaluating linear trends on increasingly long timescales would capture more of the underlying forced trend. In this context, it is a key novel result that our results show that models no longer fail the fundamental test of being able to simulate observed linear trends over the full 45-year modern satellite era.

However, we must interpret the results of the linear evaluation in the light of the recently observed abrupt decline, whereby the linear model looks increasingly less valid for observations. This again implies the emerging agreement on linear trends should not necessarily imply more confidence in model projections. From this perspective, the rapid decline provides a new context for comparing observations and models (Diamond et al., 2024) and adds evidence for which characteristics of sea ice variability the models are unable to simulate and should therefore be a focus of future studies. Therefore, while it is a tenable view that the observed rapid decline could be the first indication that the declines projected in the models could occur, there is now a need to probe the nature of this recent change, specifically the contribution of multiple timescales, and its representation in models. This will be challenging, since extremes and~~Future studies can therefore move on to more detailed model~~

~~assessments, including of~~ multidecadal variability are ~~nd rapid changes; although both are challenging~~difficult to assess due~~in the light of~~ to limited observational data~~. The rapid decline is as yet short-lived, so an improved understanding of multi-decadal sea ice variability and its representation in climate models is critical for further interpreting these results.~~ Moreover, the recent declines are still short-lived, so further years of data will add clarity to the nature of recent change. More broadly, ~~Further, processes lacking from models, such as increasing freshwater input from accelerating ice sheet melt (Swart et al., 2023), may provide further complications in the relative evolution of modelled and observed sea ice over the 21st century.~~~~T~~there are ~~of course~~ many ~~further~~ measures by which modelled sea ice may be assessed and found to have deficiencies, including seasonal and interannual variability (Zunz et al., 2013), spatial patterns (Hobbs et al., 2015), physical processes (Holmes et al., 2019), ~~,~~and relationships between trends and other variables (e.g. global warming; Rosenblum and Eisenman, 2017 or mean state, as discussed in Section 3.2~~above~~). ~~Moreover, linear trends are a limited parametric assessment of a timeseries and one could argue that the observed time series appears to display more complexity than a linear trend with noise imposed (Fig B1). However, complex behaviours, revealing the interplay of trends and variability including on long timescales, are also apparent in individual model ensemble members (Fig B1). Therefore, our argument is simply that being able to simulate linear trends is a fundamental test, and models no longer fail this fundamental test, at least over the 45-year modern satellite era.~~ ~~Future studies can therefore move on to more detailed model assessments, including of multidecadal variability and rapid changes; although both are challenging in the light of limited observational data. The rapid decline is as yet short-lived, so an improved understanding of multi-decadal sea ice variability and its representation in climate models is critical for further interpreting these results. Further, processes lacking from models, such as increasing freshwater input from accelerating ice sheet melt (Swart et al., 2023), may provide further complications in the relative evolution of modelled and observed sea ice over the 21st century.~~

~~on multiple timescales and its representation in models. , in order to critique this interpretation and ultimately assess how best to evaluate CMIP6 models.~~ ~~Our results have broad ramifications for future assessments of CMIP6 outputs. First, revising our confidence in the climate models has consequences for the attribution of historical climate changes. Secondly, we should now have some level of greater confidence in the strong projected declines in Antarctic sea ice under anthropogenic forcing (Roach et al., 2020), whereby ice becomes near-absent in summer (Holmes et al., 2022). This in turn will influence our understanding of the future evolution of all aspects of the~~Improving knowledge on the strengths and weaknesses of climate models in representing sea ice is important for understanding wider implications for Southern Hemisphere climate - including Southern Ocean heat and carbon uptake, circumpolar winds (Bracegirdle et al., 2018), and melting of the Antarctic Ice Sheet – and of marine ecosystem function; all of which underpins decisions about the mitigation of future greenhouse gas emissions and about ecosystem management.

Appendix A: Monthly trends

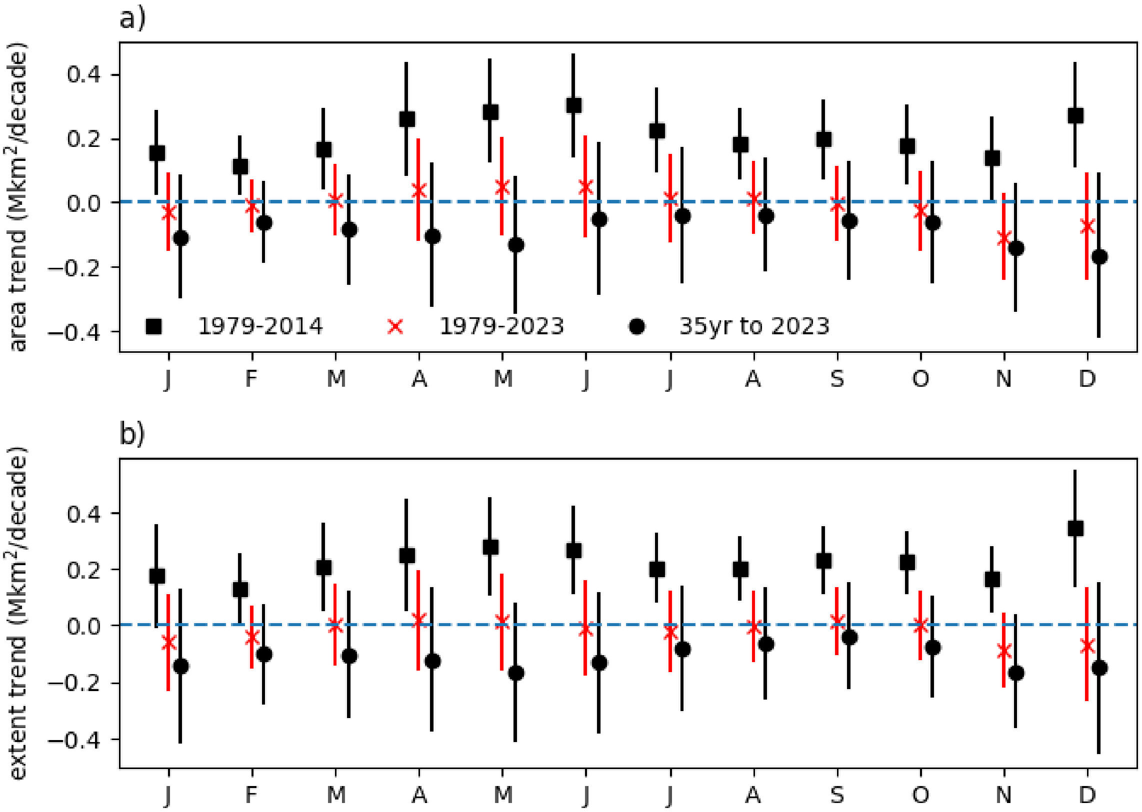

Figure A1: Observed sea ice trends in individual months for (squares) 1979-2014, (crosses) full 45-year trend 1979-2023, and (circles) 35-year trend to 2023. 1979-2023 trends are highlighted in shades of red as this period is the focus of the paper. a) Sea Ice Area, b) Sea Ice Extent. 5th-95th percentile uncertainties are indicated by vertical lines. Data are from the Sea Ice Index (see Methods).

Appendix B: CMIP6 models

	Mean Trend			Climatology		Trend piControl	n members used (available)
Model	1979-2013	1979-2023	1989-2023	February 1979-2023	Annual 1979-2023		
ACCESS_CM2	-0.049	-0.173	-0.265	0.532	7.435	-0.021	1
ACCESS_ESM1	-0.151	-0.099	-0.104	2.120	8.238	N/A	3
AWICM1	-0.405	-0.473	-0.420	1.171	9.802	0.004	1
BCC_CSM2	0.194	-0.443	-0.803	0.294	6.644	-0.027	1
CAMS_CSM1	-0.067	-0.096	-0.230	0.012	5.846	-0.023	2
CESM2	-0.369	-0.382	-0.388	1.602	8.960	-0.007	3
CESM2_WACCM	-0.474	-0.447	-0.446	1.760	9.181	-0.012	3
CIESM	-0.251	-0.261	-0.261	0.079	5.487	-0.019	1
CMCC_CM2_SR5	-0.356	-0.330	-0.328	0.679	7.568	-0.040	1
CMCC_ESM2	-0.297	-0.247	-0.254	0.719	7.699	-0.045	1
CNRM_CM6	-0.362	-0.379	-0.376	0.940	9.192	-0.018	6
CNRM_CM6_1_HR	-0.443	-0.583	-0.950	0.499	8.065	-0.065	1
CanESM5	-0.386	-0.356	-0.373	4.014	11.841	0.005	6 (19)
E3SM_1_1	-0.323	-0.360	-0.422	1.320	9.166	0.003	1
ECEarth3	-0.267	-0.222	-0.236	0.263	4.654	-0.009	6 (57)
ECEarth3_CC	-0.231	-0.126	-0.147	0.056	3.187	-0.013	1
ECEarth3_Veg	-0.149	-0.196	-0.276	0.298	4.816	-0.008	5
ECEarth3_Veg_LR	-0.325	-0.280	-0.293	0.182	4.819	-0.005	1
FGOALS_f3L	-0.122	-0.159	-0.109	0.277	6.360	N/A	1
FGOALS_g3	-0.279	-0.226	-0.135	2.214	10.813	0.000	4
FIO_ESM	-0.316	-0.342	-0.339	2.035	9.448	-0.001	3
GFDL_CM4	-0.223	-0.193	-0.159	0.529	9.791	-0.019	1
GFDL_ESM4	-0.039	-0.111	-0.075	0.641	8.455	-0.019	1
GISS_E2_1_G	-0.135	0.008	0.062	0.731	8.049	N/A	1
HadGEM3_GC31_LL	-0.514	-0.607	-0.674	1.957	8.692	N/A	3
HadGEM3_GC31_MM	-0.312	-0.313	-0.362	1.482	6.144	-0.047	4
INM_CM4_8	-0.193	-0.210	-0.228	0.242	4.386	-0.012	1
INM_CM5_0	-0.238	-0.232	-0.200	0.904	6.231	0.021	1
IPSL_CM6A_LR	-0.363	-0.384	-0.414	1.616	10.606	0.006	6
KIOST_ESM	-0.259	-0.215	-0.156	0.725	6.252	N/A	1
MIROC6	-0.014	-0.015	-0.006	0.017	1.505	-0.001	3
MIROC_ES2L	-0.072	-0.084	-0.108	0.019	1.398	0.002	6 (8)

MPI_ESM1_2_HR	-0.277	-0.274	-0.356	0.298	5.833	-0.004	2
MPI_ESM1_2_LR	-0.108	-0.078	0.019	0.259	4.325	0.000	6 (30)
MRI_ESM2	-0.325	-0.377	-0.436	2.537	11.964	-0.009	1
NESM3	-0.202	-0.283	-0.374	0.485	7.746	-0.010	2
NorESM2_LM	-0.096	-0.082	-0.102	1.385	6.238	N/A	1
NorESM2_MM	-0.014	-0.077	-0.041	1.402	6.543	N/A	1
UKESM1_0_LL	-0.721	-0.666	-0.652	2.947	9.954	0.005	5

Table B1: The models available for the study and summary values: the number of ensemble members number used (and the number available where this differs); the ensemble mean trend (Mkm^2/decade) and the climatology (Mkm^2) across the ensemble members used only for the period specified; and the trend in the pre-industrial simulation (Mkm^2/decade) . NorESM values were calculated by the authors from SIC data; all other values were obtained from the CMIP6 SIA Directory made available by the University of Hamburg and methods are fully detailed there.

Figure B1: 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. Panels are sorted by their linear trend over 1979-2023. Linear trends are shown and indicated in red (statistically significant at $p<0.05$) or grey (statistically insignificant). Each panel includes annotation showing the simulation's 1979-2023 climatology and trend. Y-axis shows SIA anomaly from 1979-2023 climatology (Mkm^2).

Appendix C: Sensitivity Tests

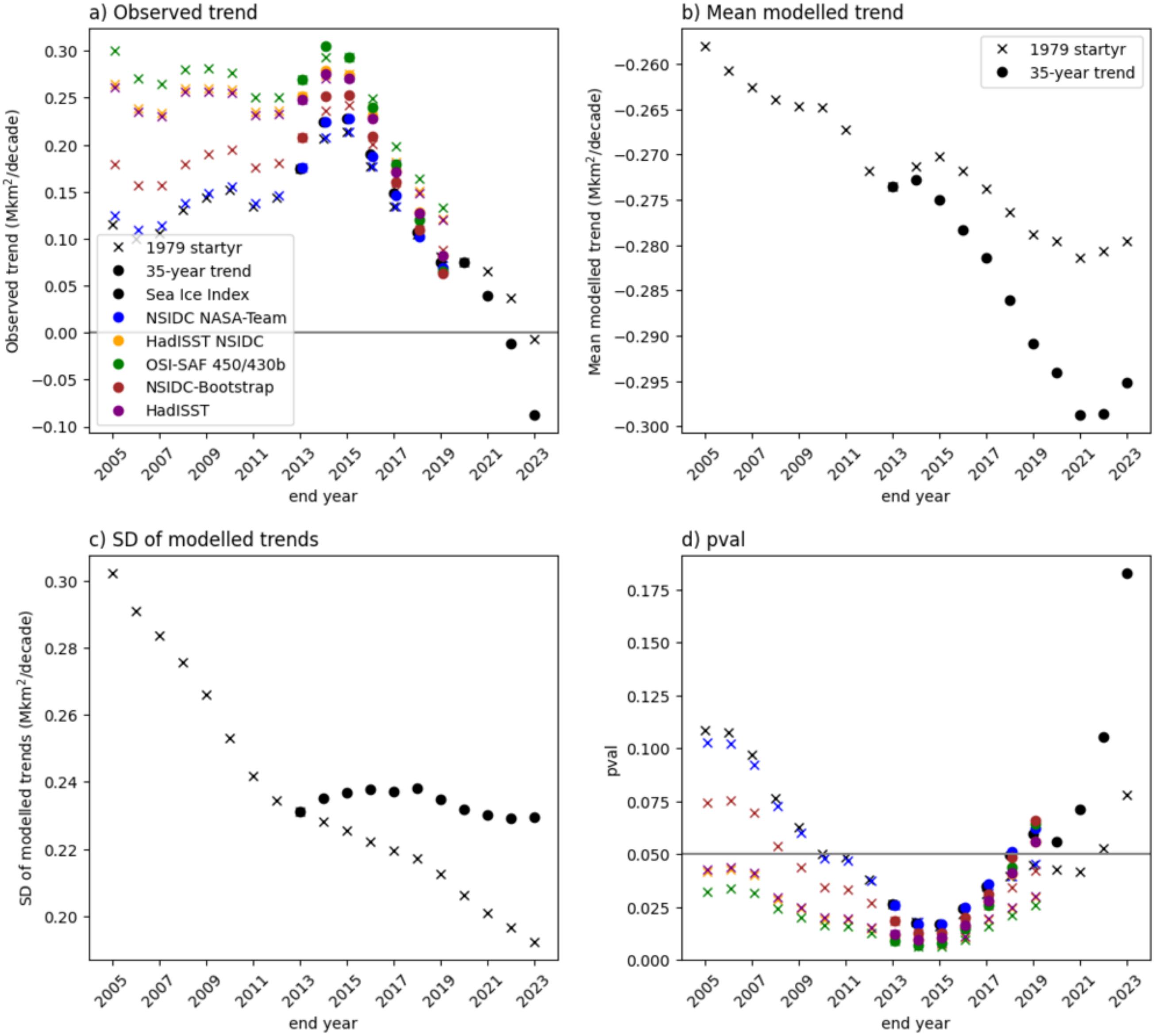

Figure C1: Contributions to the p-value shown in Figure 1d). a) observed trend; Sea Ice Index in black as in main text, other datasets as indicated. b) mean of modelled trends, c) standard deviation of modelled trends, d) p-value (as main text Figure 1d but with alternative observational estimates (Dörr, 2021)).

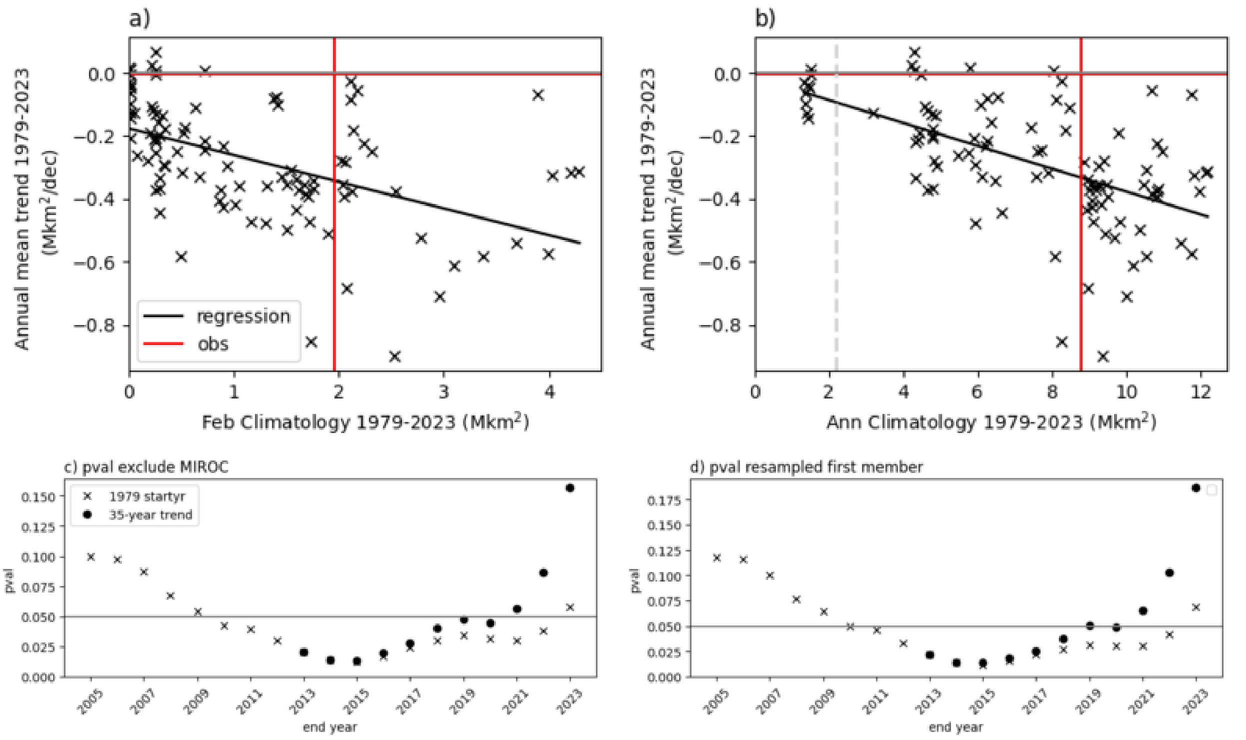

Figure C2: The role of ice-free conditions in explaining model spread, and result sensitivity to ensemble treatment. a) Scatter plot of summer (February) sea ice climatology for 1979-2023 against the annual-mean trend over 1979-2023. Maximum 6 ensemble members per model shown. b) as a) but for annual mean climatology against trend, with cutoff threshold (observed climatology/4) to exclude MIROC models indicated in grey dashed line. c) As figure 1d) but excluding MIROC models. d) As figure 1d) but using 1 random ensemble member from each model, resampled 10000 times; mean of p-values.

Sensitivity to Observational Dataset

Observational uncertainty in SIA is particularly high prior to winter 1987 (not shown) due to missing SIC data. Trends in the other datasets, in particular OSI-SAF (Figure C1, green), are in general more strongly positive than those in the Sea Ice Index (Figure C1a). Therefore, for the ‘1979 start date’ trends, these might exhibit consistency with model-simulated trends at later end dates than 2022 (Figure C1d, crosses); note that all datasets already display consistency for the 35-year trends ending in 2019 onwards (Figure C1d, dots).

Sensitivity to treatment of CMIP6 models

We also tested the sensitivity of our conclusions to our treatment of CMIP6 models. First, we tested the sensitivity to treatment of individual model ensembles. As stated in the main text, the choice of using a maximum of six ensemble members per model was to sample internal variability adequately without weighting towards models with large ensembles. By including all ensemble members (instead of a maximum of six per model), we largely add simulations from models with weak negative average trends (Table B1) and so increase consistency with observations (not shown). However, the evolution with end year of the model-observation comparison (Fig. 1d) and the broad timings of threshold crossings are unchanged. On the other hand, since curtailment to a maximum six members per model still constitutes uneven sampling across models which have different internal variabilities, we also verified that when using one ensemble member per model, results remain on average the same for 2023 end dates (Fig. C2d).

Second, we tested sensitivity to using the weaker forcing scenario ssp245 instead of ssp585 for the extension of modelled trends after 2014. The effect of forcing scenario is small early in the 21st century (Hawkins and Sutton, 2012), so that any difference arising is due to internal variability or structural differences between the models with simulations available. For the overlapping subset of 147 model-realisation combinations, ssp245 has marginally stronger trends and so is slightly less consistent with observations. In contrast, using the full ssp245 ensemble (with all available members) means including a larger ensemble of MIROC6 than in the overlapping subset or in the ssp585 ensemble; MIROC6 implausibly has virtually no sea ice year-round (Shu et al., 2020) and therefore zero trends (Holmes et al., 2022) leading to weaker mean trends and slightly greater consistency with observations. In summary, these effects are small, and so our conclusions are robust to these sensitivity tests.

Code Availability

The code for calculating trends, performing the evaluation and preparing figures is available from the corresponding author on request.

Data Availability

Sea Ice Area from the CMIP6 models is available from the University of Hamburg (UHH) CMIP6 Sea Ice Area directory (https://www.cen.uni-hamburg.de/en/icdc/data/cryosphere/cmip6-sea-ice-area.html, accessed 2023-08-17). The NSIDC Sea Ice Index v3.0 SIA (Fetterer, 2017) is available from https://nsidc.org/arcticseaicenews/sea-ice-tools/. Other observational estimates of sea ice area (Dörr, 2021) are available from https://doi.org/10.25592/uhhfdm.8559.

Author Contributions

CRH, TJB and PRH conceived the study. CRH conducted the analysis and prepared the figures. All authors discussed the results and reviewed the manuscript.

Competing Interests

The authors declare they have no conflicts of interest.

Acknowledgements

All authors received funding from NERC grant DEFIANT (NE/W004739/1); JS also received funding from Canada 150 Research Chairs program (C150 grant no. 50296). The World Climate Research Programme's (WCRP) Working Group on Coupled Modelling, which is responsible for CMIP, and the climate modelling groups, are thanked for producing and making available their model output.

References

Bracegirdle, T. J., Hyder, P., and Holmes, C. R.: CMIP5 diversity in southern westerly jet projections related to historical sea ice area: Strong link to strengthening and weak link to shift, J. Climate, 31, 195-211, 2018.

Bracegirdle, T. J., Stephenson, D. B., Turner, J., and Phillips, T.: The importance of sea ice area biases in 21st century multimodel projections of Antarctic temperature and precipitation, Geophys. Res. Lett., 42, 10,832-810,839, 2015.

Cavalieri, D., Parkinson, C., and Vinnikov, K. Y.: 30-Year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability, Geophys. Res. Lett., 30, 2003.

Diamond, R., Sime, L.C., Schroeder, D., and Holmes, C.R.: CMIP6 models rarely simulate Antarctic winter sea-ice anomalies as large as observed in 2023, Geophys. Res. Lett.,(accepted) (2024)

Dörr, J. N., Dirk; Kern, Stefan: UHH sea-ice area product, 1850-2019 (v2019_fv0.01) [dataset], https://doi.org/10.25592/uhhfdm.8559, , 2021.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, 2016.

Ferreira, D., Marshall, J., Bitz, C. M., Solomon, S., and Plumb, A.: Antarctic Ocean and sea ice response to ozone depletion: A two-time-scale problem, J. Climate, 28, 1206-1226, 2015.

Fetterer, F., K. Knowles, W. N. Meier, M. Savoie, and A. K. Windnagel.: Sea Ice Index, Version 3 Boulder, Colorado USA. National Snow and Ice Data Center. [dataset], https://doi.org/10.7265/N5K072F8, 2017.

Fogt, R. L., Sleinkofer, A. M., Raphael, M. N., and Handcock, M. S.: A regime shift in seasonal total Antarctic sea ice extent in the twentieth century, Nat. Clim. Change, 12, 54-62, 2022.

Fox-Kemper, B., H.T. Hewitt, C. Xiao, G. Aðalgeirsdóttir, S.S. Drijfhout, T.L. Edwards, N.R. Golledge, M. Hemer, R.E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I.S. Nurhati, L. Ruiz, J.-B. Sallée, A.B.A. Slangen Y. Yu: Ocean, Cryosphere and Sea Level Change, 1211-1362, 10.1017/9781009157896.011., 2021.

Gagné, M. È., Gillett, N., and Fyfe, J.: Observed and simulated changes in Antarctic sea ice extent over the past 50 years, Geophys. Res. Lett., 42, 90-95, 2015.

Gallaher, D. W., Campbell, G. G., and Meier, W. N.: Anomalous variability in Antarctic sea ice extents during the 1960s with the use of Nimbus data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7, 881-887, 2013.

Gupta, A. S., Jourdain, N. C., Brown, J. N., and Monselesan, D.: Climate Drift in the CMIP5 Models, J. Climate, 26, 8597-8615, https://doi.org/10.1175/JCLI-D-12-00521.1, 2013.

Hawkins, E. and Sutton, R.: Time of emergence of climate signals, Geophys. Res. Lett., 39, 2012.

Hobbs, W. R., Bindoff, N. L., and Raphael, M. N.: New Perspectives on Observed and Simulated Antarctic Sea Ice Extent Trends Using Optimal Fingerprinting Techniques, J. Climate, 28, 1543-1560, https://doi.org/10.1175/JCLI-D-14-00367.1, 2015.

Holmes, C., Bracegirdle, T., and Holland, P.: Antarctic sea ice projections constrained by historical ice cover and future global temperature change, Geophys. Res. Lett., 49, e2021GL097413, 2022.

Holmes, C. R., Holland, P. R., and Bracegirdle, T. J.: Compensating biases and a noteworthy success in the CMIP5 representation of Antarctic sea ice processes, Geophys. Res. Lett., 46, 4299-4307, 2019.

Kostov, Y., Marshall, J., Hausmann, U., Armour, K. C., Ferreira, D., and Holland, M. M.: Fast and slow responses of Southern Ocean sea surface temperature to SAM in coupled climate models, Clim. Dynam., 48, 1595-1609, 2017.

National Academies of Sciences, E. and Medicine: Antarctic sea ice variability in the southern ocean-climate system: Proceedings of a workshop, 2017.

Meier, W. N., Gallaher, D., and Campbell, G. G.: New estimates of Arctic and Antarctic sea ice extent during September 1964 from recovered Nimbus I satellite imagery, Cryosphere, 7, 699-705, 10.5194/tc-7-699-2013, 2013.

Notz, D.: Sea-ice extent and its trend provide limited metrics of model performance, Cryosphere, 8, 229-243, 2014.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., and Lowe, J.: The scenario model intercomparison project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9, 3461-3482, 2016.

Parkinson, C. L.: A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic, Proceedings of the National Academy of Sciences, 116, 14414-14423, 2019.

Polvani, L. M. and Smith, K. L.: Can natural variability explain observed Antarctic sea ice trends? New modeling evidence from CMIP5, Geophys. Res. Lett., 40, 3195-3199, https://doi.org/10.1002/grl.50578, 2013.

Purich, A. and Doddridge, E. W.: Record low Antarctic sea ice coverage indicates a new sea ice state, Communications Earth & Environment, 4, 314, 2023.

Roach, L. A., Dörr, J., Holmes, C. R., Massonnet, F., Blockley, E. W., Notz, D., Rackow, T., Raphael, M. N., O'Farrell, S. P., Bailey, D. A., and Bitz, C. M.: Antarctic Sea Ice Area in CMIP6, Geophys. Res. Lett., 47, e2019GL086729, https://doi.org/10.1029/2019GL086729, 2020.

Rosenblum, E. and Eisenman, I.: Sea ice trends in climate models only accurate in runs with biased global warming, J. Climate, 30, 6265-6278, 2017.

Schlosser, E., Haumann, F. A., and Raphael, M. N.: Atmospheric influences on the anomalous 2016 Antarctic sea ice decay, Cryosphere, 12, 1103-1119, 2018.

Schneider, D. P. and Deser, C.: Tropically driven and externally forced patterns of Antarctic sea ice change: Reconciling observed and modeled trends, Clim. Dynam., 50, 4599-4618, 2018.

Schroeter, S., O'Kane, T. J., and Sandery, P. A.: Antarctic sea ice regime shift associated with decreasing zonal symmetry in the Southern Annular Mode, Cryosphere, 17, 701-717, 2023.

Seviour, W., Codron, F., Doddridge, E. W., Ferreira, D., Gnanadesikan, A., Kelley, M., Kostov, Y., Marshall, J., Polvani, L., and Thomas, J.: The Southern Ocean sea surface temperature response to ozone depletion: A multimodel comparison, J. Climate, 32, 5107-5121, 2019.

Shu, Q., Wang, Q., Song, Z., Qiao, F., Zhao, J., Chu, M., and Li, X.: Assessment of sea ice extent in CMIP6 with comparison to observations and CMIP5, Geophys. Res. Lett., 47, e2020GL087965, 2020.

Siegert, M. J., Bentley, M. J., Atkinson, A., Bracegirdle, T. J., Convey, P., Davies, B., Downie, R., Hogg, A. E., Holmes, C., and Hughes, K. A.: Antarctic extreme events, Frontiers in Environmental Science, 11, 1229283, 2023.

Swart, N., Martin, T., Beadling, R., Chen, J.-J., England, M. H., Farneti, R., Griffies, S. M., Hatterman, T., Haumann, F. A., and Li, Q.: The Southern Ocean Freshwater release model experiments Initiative (SOFIA): Scientific objectives and experimental design, EGUsphere, 2023, 1-30, 2023.

Turner, J. and Comiso, J.: Solve Antarctica’s sea-ice puzzle, Nature, 547, 275-277, 2017.

Turner, J., Phillips, T., Marshall, G. J., Hosking, J. S., Pope, J. O., Bracegirdle, T. J., and Deb, P.: Unprecedented springtime retreat of Antarctic sea ice in 2016, Geophys. Res. Lett., 44, 6868-6875, 2017.

Zhang, L., Delworth, T. L., Cooke, W., and Yang, X.: Natural variability of Southern Ocean convection as a driver of observed climate trends, Nat. Clim. Change, 9, 59-65, 2019.

Zhang, L., Delworth, T. L., Yang, X., Zeng, F., Lu, F., Morioka, Y., and Bushuk, M.: The relative role of the subsurface Southern Ocean in driving negative Antarctic Sea ice extent anomalies in 2016–2021, Communications Earth & Environment, 3, 302, 2022.

Zunz, V., Goosse, H., and Massonnet, F.: How does internal variability influence the ability of CMIP5 models to reproduce the recent trend in Southern Ocean sea ice extent?, Cryosphere, 7, 451-468, 2013.